# Lower Passaic River Study Area Remedial Investigation and Feasibility Study
# EPA Review of the Cooperating Parties Group Approach to Mapping Contaminants of Potential Concern

## Executive Summary


The CPG's draft Remedial Investigation (RI) Report (February 2015) includes a description of the approach used for estimating COPC concentrations in sediment in locations where data were not collected. The CPG's COPC mapping approach (hereafter referred to as the Mapping Approach) provides a basis (i) for assigning initial conditions in the contaminant fate and transport model and (ii) for developing remedial alternatives. With regards to developing remedial alternatives, the CPG used results of the mapping to estimate changes in sediment surface weighted average concentrations (SWACs) resulting from targeting areas for remediation where estimated concentrations exceed a threshold concentration referred to as a remedial action limit (RAL). The CPG evaluated post-remedial SWACs for a range of RALs to develop a SWAC vs. RAL relationship. EPA is concerned with the reliability of the relationship developed by the CPG, which is directly tied to the reliability of the Mapping Approach.

The Mapping Approach involves delineating zones throughout the river based on physical features, including water depth (i.e. channel versus left and right shoal), bed composition (silt and non-silt), and from RM 2.3 and RM 7.8, bathymetric changes in the channel over two time periods (1949-2011 and 1966 – 2011). Contaminant data are used to assign concentrations to unsampled areas of the same type zone using Thiessen polygons (i.e. areas of constant concentration surrounding each data point) to distribute the data. Thiessen polygons can result in sharp differences in concentration between areas of uniform concentration, in spite of the well-documented heterogeneity of chemical concentrations in sediment at distances smaller than the size of the polygons. EPA has concerns with the accuracy of the Mapping Approach for predicting appropriate remedial areas for several reasons, including the following:

- Developing estimates of concentrations in discontiguous unsampled areas of the same zone type results in assignment of concentrations from individual samples beyond reasonable distances, including locations separated by areas of a different zone type when other samples are otherwise available nearby.
- The erosion/deposition thresholds used to delineate the sub-areas from RM 2.3 to RM 7.8 were established based on the contaminant concentration data. Basing the definition of the zones on the concentration data and then making predictions of concentrations in unsampled areas of the same zone leads to unreliable predictions.
- Bathymetry changes are not reliably accurate to the resolution of the erosion/deposition thresholds (0.4 and 0.5 feet) used for the zone delineation.
- The approach is highly sensitive to small changes in the data and is therefore not a robust predictive model. For example, the analysis herein shows that the concentration of a particular


620522

polygon can increase by 2 orders of magnitude with the addition of one new data point. This sensitivity to new data or changes in data handling decisions is characteristic of models that may not generalize well, and suggests that there may be inaccuracies in the SWAC vs. RAL relationship developed based on the Mapping Approach.

- The documentation for the Mapping Approach did not include an evaluation of uncertainty or margin of error in the estimates of concentrations in unsampled areas or a demonstration that the approach improves the prediction of concentrations at unsampled areas over alternative, generally accepted approaches.

To assess the reliability of the Mapping Approach EPA performed the following analyses:

- Evaluation of the sensitivity of the estimated concentrations to the addition of new data and changes in the mapping rules, which have evolved over time,
- Evaluation of the reliability of a SWAC versus RAL relationship developed using the Mapping Approach. This included analyses using RI and pre-design data from the area near where the RM 10.9 removal action was completed, and a simulation study in which the true SWAC versus RAL relationship is known.

These analyses were conducted using a 2,3,7,8-TCDD RAL of 500 ppt (ng/kg) and a post-remedial SWAC of 150 ppt, for illustrative purposes. However, these values are used for analysis and discussion of the approach only; their inclusion in this document should not be interpreted as EPA endorsing or implicitly accepting the appropriateness of these specific values.

EPA's concern about the Mapping Approach and resulting SWAC versus RAL relationship is that it can lead to overly optimistic assessments of the volume, schedule and cost of remediation required to achieve a needed risk reduction.

**Lower Passaic River Study Area Remedial Investigation and Feasibility Study**
**EPA Review of the Cooperating Parties Group Approach to**
**Mapping Contaminants of Potential Concern**

# 1 Introduction to COPC Mapping Issue

The Cooperating Parties Group's (CPG's) draft Remedial Investigation (RI) report dated February 2015 includes an approach to mapping Contaminants of Potential Concern (COPCs) that relies on correlation between COPC concentrations and physical characteristics of the sediment, hereafter referred to as the Mapping Approach. As presented by the CPG, the Mapping Approach results in the conclusion that targeting a remedial action limit (RAL) of 500 ng/kg for 2,3,7,8-TCDD from River Mile (RM) 0 to RM 14 of the 17-mile Lower Passaic River Study Area (LPRSA) will result in a post-remedial surface weighted average concentration (SWAC) for 2,3,7,8-TCDD of 150 ng/kg.

While EPA recognizes that, in general, patterns in contaminant fate and transport can be related to geomorphic conditions, physical properties of sediment, and hydrodynamic forces within rivers, EPA concludes that the Mapping Approach does not predict sediment concentrations reliably enough to be used to develop alternatives in the Feasibility Study (FS). The Mapping Approach hinges on the accuracy of the COPC mapping strategy used, which was largely developed retrospectively. In contrast to prospective studies where hypotheses are stated a priori and tested with subsequent data, retrospective analyses seek to find patterns in predictor variables that are consistent with the previously measured dependent variables—effectively looking for causes to match effects. Models developed retrospectively commonly do not predict new conditions as well as they predict, or appear to predict, the samples used to develop them. This phenomenon is known as "over fitting" (Harrell, 2001).

This paper evaluates two primary issues; 1) the reliability of the Mapping Approach, and 2) the accuracy of the forecast that a 500 ng/kg RAL will achieve a post-remedial SWAC of 150 ng/kg, notwithstanding the reliability of the mapping method. Please note that while this paper does evaluate the accuracy of the SWAC vs. RAL relationship proposed by the CPG, it does not in any way endorse or implicitly accept that these are appropriate target values for the Lower Passaic River Study Area (LPRSA) RI/FS.

# 2 Problem Statement

The CPG proposes that a remedial action limit of 500 ng/kg could achieve a post-remedial average 2,3,7,8-TCDD concentration of approximately 150 ng/kg (83% reduction) while remediating 145 acres (16%) of the surface area of the lower 14 miles of the River (for data included in this analysis, concentrations of 2,3,7,8-TCDD above RM 14 were all found to be below 500 ng/kg)[1]. The implications of this assertion can be illustrated by considering the mathematical relationship among the percentage

[1] These numerical relationships are approximate based on information from the CPG which varies somewhat depending on the mapping version and the river miles under consideration. Evaluations conducted in this report are not sensitive to small changes in these specifics from one version to another.

reduction in SWAC, the remedial footprint size as a percentage of overall surface area, the ratio of target to non-target areas, and assumptions about post-remedial concentrations (Figure 1). By reorganizing terms in the defining equation for SWAC at the upper right corner of Figure 1, it can be shown that for fixed ratios of residual to pre-remedial concentrations $\left(\frac{C_{Residual}}{C_R}\right)$ and fixed ratios of remediated to unremediated concentrations$\left(\frac{C_R}{C_U}\right)$, the relative reduction in SWAC is a function of the proportion of area remediated, with proportional SWAC reduction increasing to 100% as the proportion of the footprint increases.

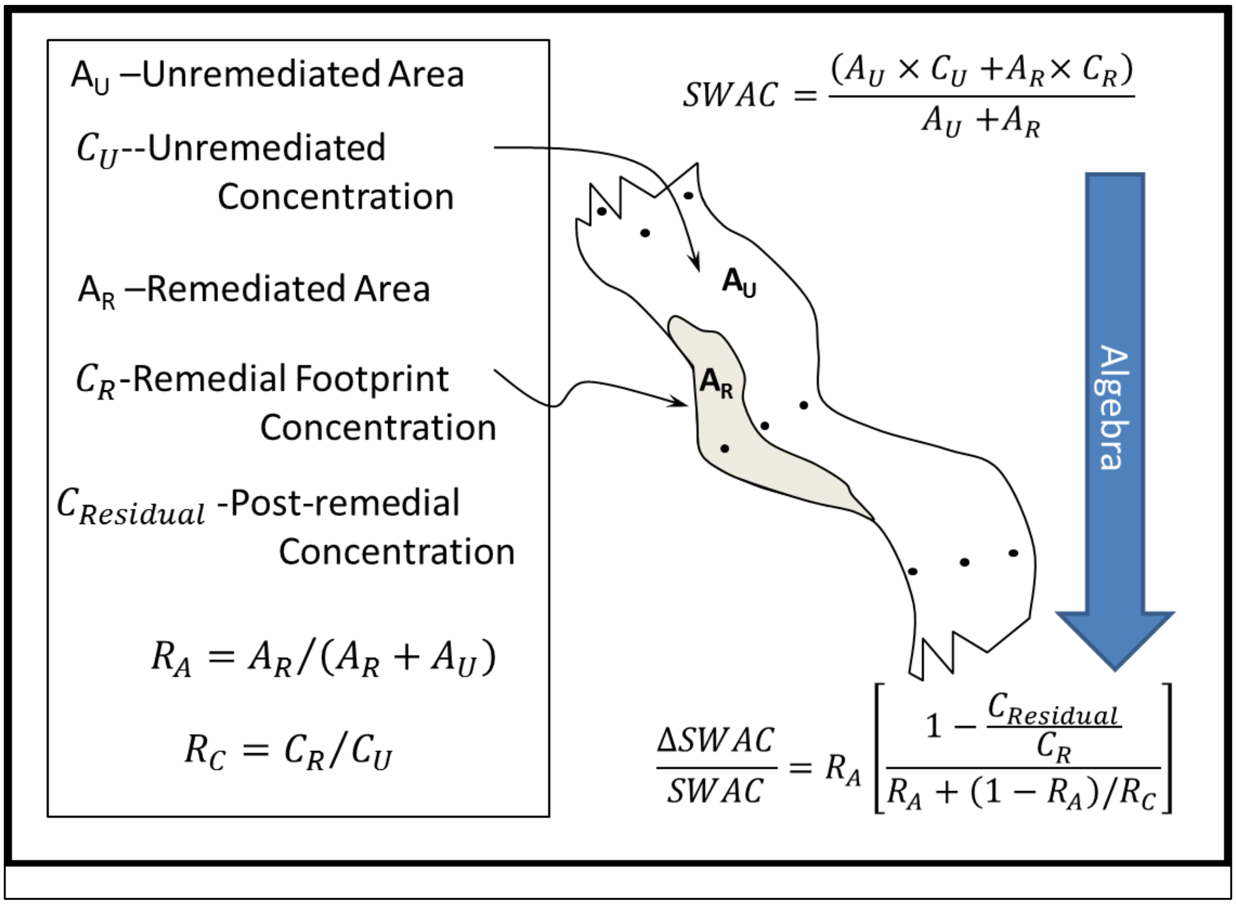


Figure 1. Mathematical relationships constraining reduction in SWAC associated with remedial footprint.

The concentration terms in this relationship represent spatial averages reflecting the average contaminant concentrations of the areas that would be remediated at the scale of remedial actions, as opposed to the scale of individual samples. Actual performance will be determined by the average of concentrations found within remedial and non-remedial footprints. This point is emphasized here because the Mapping Approach is based on calculations using individual samples as surrogates for larger areas identified on maps as Thiessen polygons. Accurately forecasting effects of remedial alternatives requires accurate estimates of these concentration terms.

The relationship described in Figure 1 is plotted in Figure 2, with separate curves representing target to non-target concentration ratios ($R_C$) of 2, 4, 10, 100 and 1000 showing that as the ratio of target to non-target concentrations increases, the percentage area necessary to achieve a specified level of reduction decreases. For example, to achieve a 60% reduction in SWAC when the concentration ratio is 2 to 1, more than 40% of the surface area would need to be remediated. To achieve a similar level of effectiveness with a smaller remedial footprint, say 15% of the surface area, would require a concentration ratio of approximately 10 to 1. Achieving higher concentration ratios in practice requires more highly resolved (i.e. more accurate) deposit delineation.

This plot is general and can be used to evaluate any proposed remedial alternative. For example, the CPG has stated that its proposed RAL of 500 ng/kg would achieve a reduction in SWAC of approximately 84% while remediating approximately 16% of the surface area. EPA recalculated these values based on the most recent mapping provided to EPA, and found that the proposed RAL of 500 ng/kg would be forecast to achieve an 81% reduction in SWAC with a remedial footprint of 14.5% of the LPRSA. Locating these percentages on Figure 2 indicates that to achieve this result, the target to non-target concentration ratio must be approximately 26 to 1.

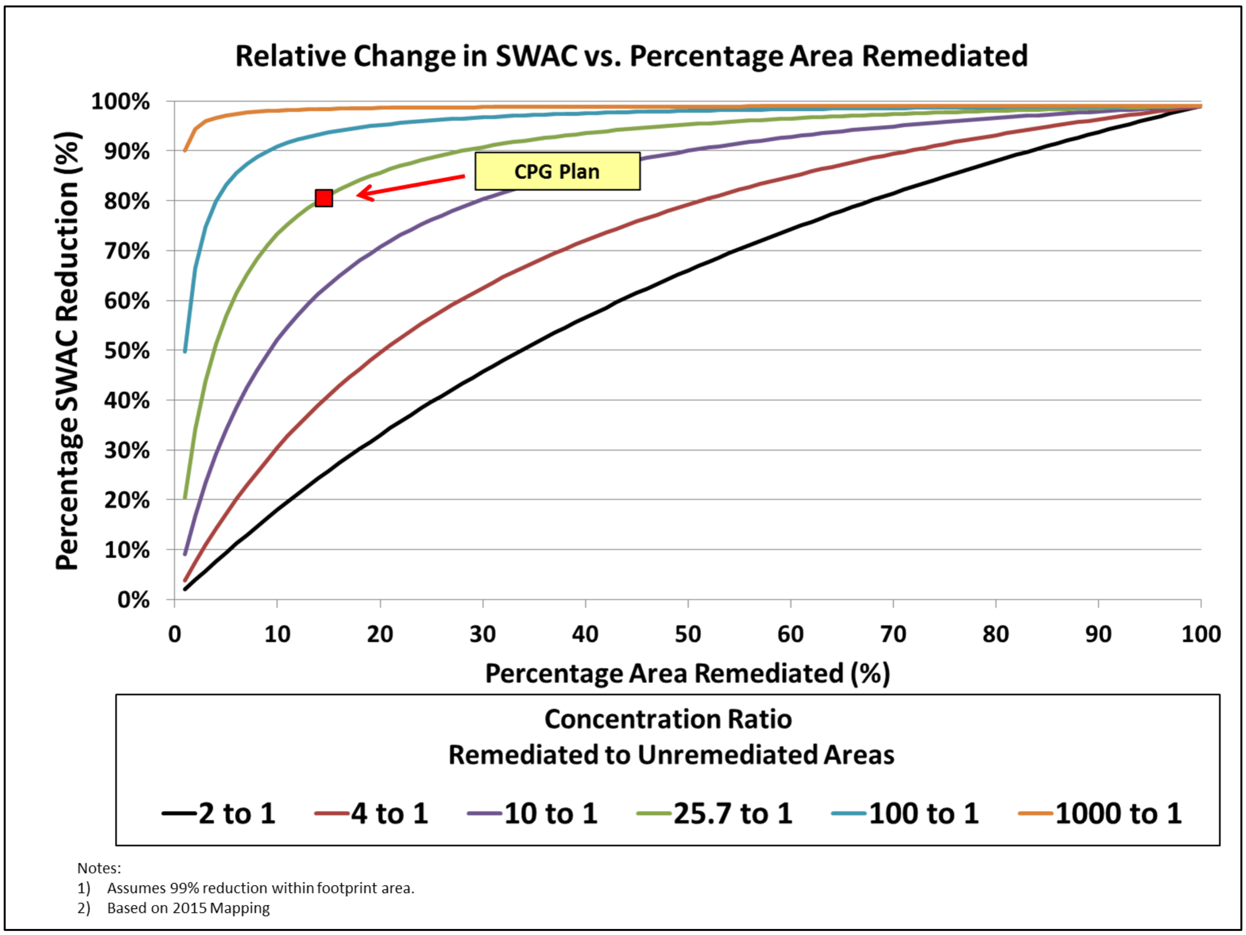


Figure 2. Percentage reduction in SWAC plotted against percentage surface area remediated by ratio of average concentration in remediated areas to that in the un-remediated areas.

The relationship described in Figure 2 is important because these generally applicable mathematical relationships impose performance constraints on remedial implementation, without which the forecast reduction in SWAC (i.e. risk reduction) might not be attained in practice. The CPG's proposed remedial alternative would require a relatively high target to non-target concentration ratio of 26 to 1 to be successful, implying that delineation accuracy must be more highly resolved than if a higher percentage area were to be remediated. Conversely, less highly resolved mapping accuracy would be required for a remedy based on a concentration ratio of 2 to 1. The ability to accurately predict and reach a successful outcome would ultimately be determined by the accuracy of the COPC mapping procedure. In the following sections, the Mapping Approach and the calculations supporting the relationship between 500 ng/kg RAL and 150 ng/kg PRG (post-remedial SWAC) are evaluated within the context of this constraining relationship between SWAC and remedial footprint.

The Mapping Approach is used to develop a forecast of the effectiveness of a RAL in achieving a post-remedial SWAC. The accuracy of the forecast requires unbiased estimation of the average concentration within the two delineated areas—target and non-target. Misclassification of sediments relative to the RAL, i.e., either understatement of the non-target average or overstatement of the target average, would cause overstatement of relative reduction in SWAC. While these potential biases are not unique to the approach utilized by the CPG for the LPRSA, the errors in this case may be unusually large due to assumed sharp divisions between target and non-target areas due to potential over fitting, and because of demanding delineation accuracy requirements associated with the relatively small proposed remedial footprint.

# 3 Overview of the CPG's COPC Mapping Approach

The Mapping Approach is based on assigning samples to groups, within which areas on the maps are assigned concentrations based on the nearest sample with a like group assignment. Surface 2,3,7,8 TCDD groupings were defined by breaking the river into gross geomorphic groups for the left shoal, channel and right shoal. The channel was separated into five groups. Below RM 2.3 is treated as one group, between RM 2.3 and RM 7.8 is separated into three groups based upon bathymetric changes since 1949 and 1966, and above RM 7.8 is the final group. The river upstream of RM 7.8 is further subdivided into shoals and silt deposits based on grain size data. The silt deposits include areas in both the shoals and channel.

These groups forming the basis of the Mapping Approach were developed through a retrospective procedure wherein group definitions were developed based on graphical observation of patterns in 2,3,7,8-TCDD concentrations which were used to set erosional thresholds and geographical boundaries. Subsequently, these retrospectively defined groups were used for prediction at unsampled locations (i.e. for interpolation).

The procedure as it is understood included the following steps.

1) Subdivide the LPR between RM 0-7.8 and RM 7.8-17.4
2) For RM 7.8 to RM 17.4, define general stratification of the river into:

a. Left shoal (36 acres)
b. Right shoal (33 acres)
c. Channel (228 acres)
d. Silt deposits which may occur in either shoals or channel areas (44 acres)

3) For RM 0 to 7.8 define general stratification of the river as for RM 7.8 to RM 17.4:
    a. Left Shoal (104 acres)
    b. Right Shoal (270 acres)
    c. Channel below RM 2.3 (110 acres)
    d. Channel between RM 2.3 and 7.8 (181 acres)
4) Overlay analytical sampling locations and assign changes in bathymetric surface elevations
5) Create scatter plot of concentration data against erosion and set erosion thresholds based on graphical concentration patterns:
    a. "Group 2" --non-depositional since 1949 (19 acres)
    b. "Group 3" --moderately depositional with variable 2,2,7,8-TCDD (50 acres)
    c. "Group 4" --highly depositional with less variable 2,3,7,8-TCDD (112 acres)
6) Generate Thiessen polygons for each set of locations separately by group.
7) Clip Thiessen polygons at geographic stratum boundaries.
8) Assign contaminant concentrations to each Thiessen polygon based on the sample(s) within each group. There is only one sample per Thiessen polygon, and in some cases Thiessen polygons are subdivided into geographically discontiguous areas with intervening spaces being assigned to other groups. In these cases a single sample may be used to assign concentrations to several distinct geographical areas on the maps, despite the availability of other samples in closer proximity. .
9) Make additional adjustments based on professional judgment.

The Mapping Approach was used to predict the current concentrations of 2,3,7,8-TCDD throughout the 14 miles.

## 4 Evaluation of Mapping Method Development

Strong spatial relationships predictive of concentrations exceeding certain thresholds could present opportunities to inform the remedy selection process and ultimately the remedial design. However, for such relationships to be useful, they need to satisfy at least three threshold criteria.

1) The relationship needs to be predictive of concentrations in unsampled areas.
2) The relationship needs to reliably differentiate areas where average surface concentrations exceed risk-based thresholds from those that do not.
3) For Groups 2, 3, and 4 (identified above), bathymetric change trends need to be predictable and consistent in time for a given location.

The information provided by the CPG in presentations, the draft RI report and memoranda (AQEA, 2013; Connolly, 2014; and Cooperating Parties Group, 2015) indicates that the methods used to develop the relationship may not provide a reliable basis for differentiating areas exceeding certain thresholds from those not exceeding the thresholds.

In the proposed predictive model, statistical hypotheses are represented by the selection of interpolation group boundaries defined by erosion cutoff values (for example -0.4 feet of change since 1949 is such a cutoff or threshold value). As is noted in Footnote 8 of Appendix J of the draft RI Report, the thresholds defining data groups were selected retrospectively based on inspection of graphs of 2,3,7,8-TCDD concentration plotted against bathymetric change:

> *"The cutoff of 0.4 feet of erosion for defining interpolation group boundaries was selected based on the transition in 2,3,7,8-TCDD concentrations observed at approximately this value."*

In selecting these threshold values, chemical concentration, the dependent variable to be predicted, was used to define the groups which will ultimately be used to make the predictions necessary for mapping COPCs at unsampled locations. The fundamental weakness with this approach was identified by Smith (Smith 1983) who developed a mathematical proof showing that retrospective use of the dependent variable to define group membership would lead to inaccurate statistical inference (i.e. unreliable model predictions due to over fitting).

## 4.1 Erosion Thresholds

Figures 3 through 6 provide annotated reproductions of graphics from materials provided by the CPG to EPA. They show that key thresholds vary and mapping groups are inconsistently defined across the presentations. In each of these figures, it can be seen that nearly all samples with some degree of deposition exceed 100 ng/kg, irrespective of the groups used, suggesting that the groups identified as providing meaningful changes in concentration distributions are unlikely to be useful for comparing remedial alternatives if the PRGs to be attained are below 100 ng/kg. Horizontal red lines representing RALs of 500 ng/kg (solid red), 300 ng/kg (black dashed) and 100 ng/kg (heavy red dashed) have been added to Figure 3, which was originally provided to EPA in 2013 (AQEA 2013). It can be seen that in Group 2, no samples exceed 500 ng/kg and in Groups 3 and 4, only a small percentage of samples exceed 500 ng/kg. Using these Groups, a relatively small remedial footprint was identified by the CPG for remediation, based on the assumption that a remedial design would focus on removal of the apparent extremes identified.

With a remedial action limit slightly less than 500 ng/kg this same predictive model and application would identify a much larger footprint for remediation, because the majority of samples are between 100 and 500 in all but Group 2. Also, Group 2 is relatively small in comparison to Groups 3 and 4, which constitute the majority of the lower 8 miles of the river.

The following evaluations consider the CPG's predictive model and issues related to its development with particular focus on uncertainties related to the selection of erosion thresholds.

### 4.1.1 Uncertainty in Erosion Thresholds

Figure 3 shows the first set of data groups presented by the CPG as evidence that low concentrations are predicted by areas exhibiting net erosion between 1949 and 2010. The erosion threshold for this group is presented as a single vertical line at 0.0 and is treated as a bright line differentiating Groups 2 and 3 in the predictive model. Previous analyses (Kern et al 2009) have shown that uncertainty in bathymetric change for modern single beam surveys is generally no better than plus or minus 6 inches and may be much less certain between survey lines. Bathymetric change comparisons involving older surveys, such as the 1949 survey, may be even more uncertain. To show that uncertainty, additional vertical dashed lines have been added in red to Figure 4. These uncertainty bounds of approximately plus or minus 6 inches show that even modest amounts of uncertainty around the 0.0 foot threshold or the 4 foot threshold would change membership of samples among groups and effectively eliminate apparent differences in mean concentrations among Groups 2, 3 and 4.

### 4.1.2 Threshold Selection Lacking Scientific Underpinnings

The proposed threshold values were not selected for physics-based scientific reasons, which suggests that their reliability for making predictions may be low. Conceptually, the expectation of lower concentrations in areas without sediment accumulation since 1949, and high concentrations in areas where sediment accumulation ceased around the mid 1960's follows logically. The concern is in the selection of the specific thresholds, which are not scientifically based, but rather are assigned to separate individual data points without consideration of the reliability of concentration estimates in unsampled areas.

Figure 3 shows that in their first presentation of the method to EPA, the CPG defined erosional and depositional areas based on comparisons between 1949 and 2010 bathymetric surveys and identified 0, 4 and 6 feet as thresholds defining Groups 2, 3 and 4, conditional on the particular river miles under consideration. According to Section 1.2.1 of Appendix J of the draft RI report, the threshold values were derived by examining the analytical data to identify the erosional/depositional intervals associated with transitions in 2,3,7,8-TCDD concentrations. Inspection of Figure 3 reveals that just a small number of high concentrations within the depositional window defining Group 3 drive this perception. The 6 or 7 highest measured concentrations were found for deposition values between 0 and 4 feet.

Given that the greatest deposition occurs in the lower 8 miles, and that a major source of 2,3,7,8-TCDD was located at approximately RM 3.5, it is possible that these few extreme values are actually a product of proximity to the source, grain size, and/or organic carbon variation, for which these erosional groupings may be a surrogate. It is also possible that these few extreme samples are a product of a biased sampling approach.

The only thing known about these thresholds is that they were selected to match anecdotal characteristics of the sample data in hand. As such, there is no evidence, physical or statistical, that these erosion/deposition threshold values could be expected to provide reliable predictions at unsampled locations. This is an example of the problem of over-fitting described above.

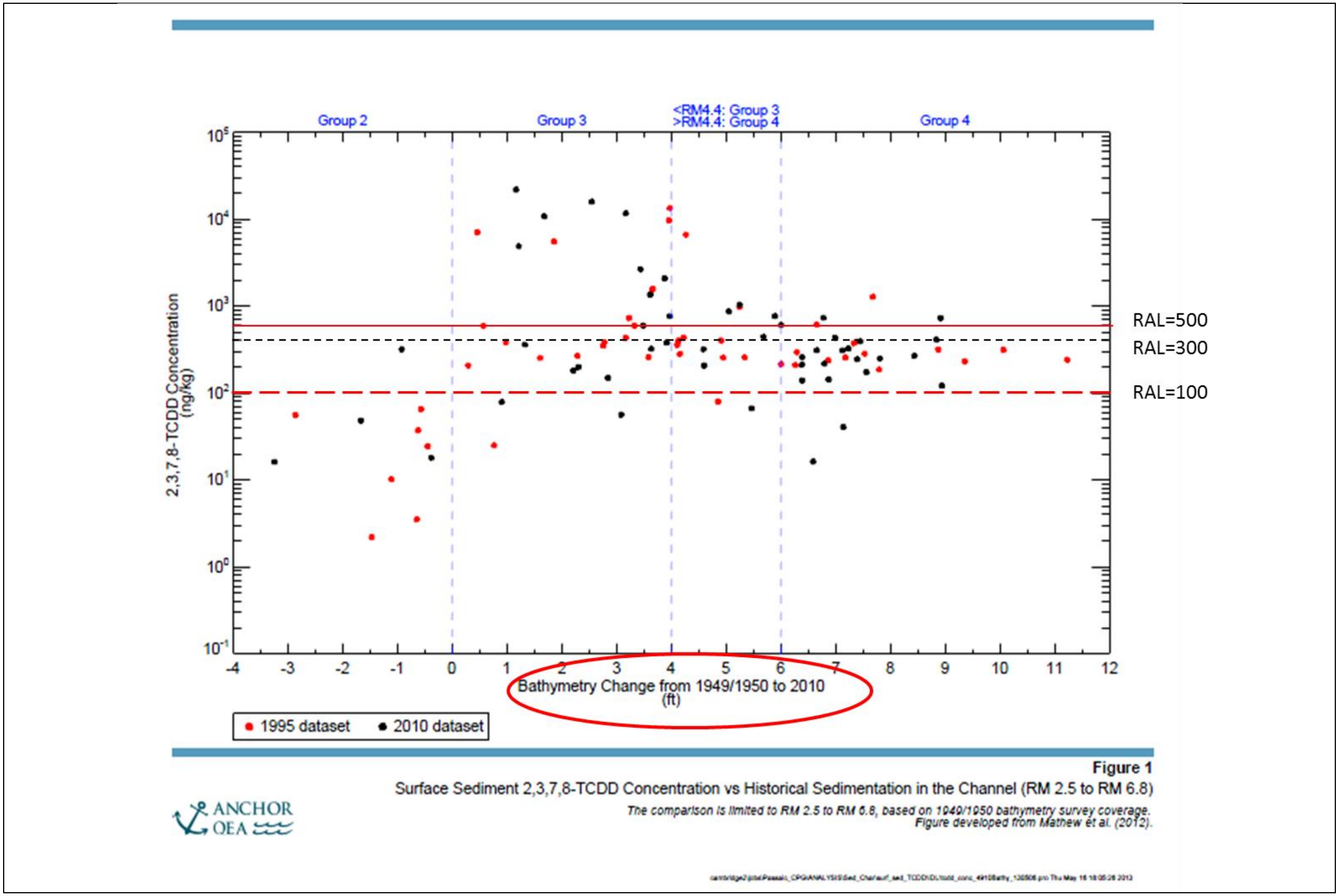


Figure 3. Surface sediment 2,3,7,8-TCDD concentration segregated by CPG erosional groupings and plotted against bathymetric change between 1949 and 2010.

### 4.1.3 Threshold Definitions Vary Among Presentations

Key threshold values defining predictive erosional and depositional groups vary substantially from one version of the mapping approach presented by CPG to the next. This variation in threshold values seems to be designed to maintain separation between high and low contaminant concentrations among groups which were initially defined based on the analytical data (again, see Footnote 8, Appendix J of the Draft RI). These evolving versions of the mapping result in substantively differing interpretations and meaningful differences in mapped areas, and illustrate the issues described above related to uncertainty in bathymetric change data and the lack of scientific underpinnings of the thresholds. When data or mapping rules change, the empirical thresholds must also change to maintain analytical data groups defined previously.

The first definition of long term net erosion or deposition, which appeared in the 2013 CPG Mapping Memorandum, was based on a comparison of 1949 and 2010 bathymetric surfaces, as shown in Figure 3. Then, a revised definition appeared in Appendix J of the draft RI that was based on a comparison of 1949 and 2011 bathymetric surfaces, as shown in Figure 4. With this change in bathymetric surfaces under comparison, the threshold also changed from 0.0 feet to -0.4 feet. Had the threshold been maintained at 0.0 feet in the second comparison, two samples exceeding 5,000 ng/kg (circled samples in Figure 4) would have been classified as belonging to the non-depositional group—contrary to the hypothesis put forth in the 2013 CPG Mapping Memo that this group only includes concentrations of 100 ng/kg or less. In the second comparison, changing the threshold to -0.4 still resulted in sample 12A-0427—the fourth highest 2,3,7,8-TCDD concentration in the data set—being classified as part of the non-depositional group (Group 2), which was still contrary to apparent group differences described in the first report. However, in Appendix J of the draft RI report, the geographic coordinates of this sample were revised using data handling techniques unique to this location, and as a result, this sample became associated with a location approximately 5 feet (1 pixel on the interpolated bathymetric surface map) away and was thus reassigned to Group 3. Figure 5 shows the new location just to the right of the -0.4 foot threshold, and Figure 6 provides a reproduction of the map from the draft RI showing the specialized data handling technique that was tailored to sample 12A-0427.

Given the uncertainty in the estimates of erosion and deposition based on the bathymetric surfaces, it is difficult to establish with any confidence the group to which this particular sample—or any of the samples ascribed to deposition interval from -0.5 to 1 foot—can reliably be assigned. Given that the definition of Group 2 (non-depositional since 1949) is based on a small number of samples, it is unlikely that the derived relationship will in fact be predictive of concentrations elsewhere within Group 2. Simply moving the erosion threshold back to 0.0 feet would in fact result in a range of 4 orders of magnitude within the non-depositional group, which would exceed the variation seen currently in Groups 3 or 4.

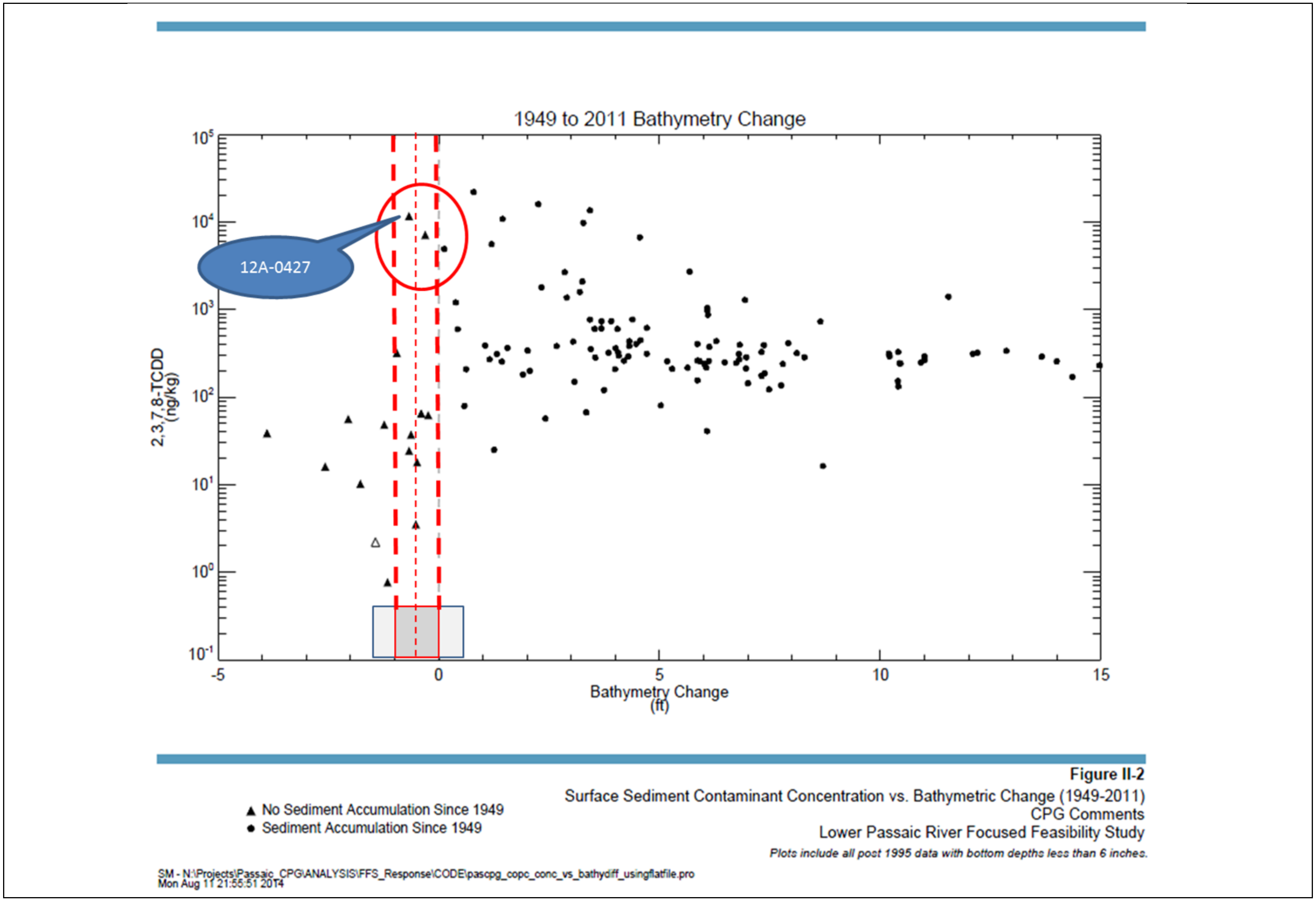

Figure 4. Erosion threshold based on comparison of erosion between 1949 and 2011 with uncertainty bands overlain on original graphic.

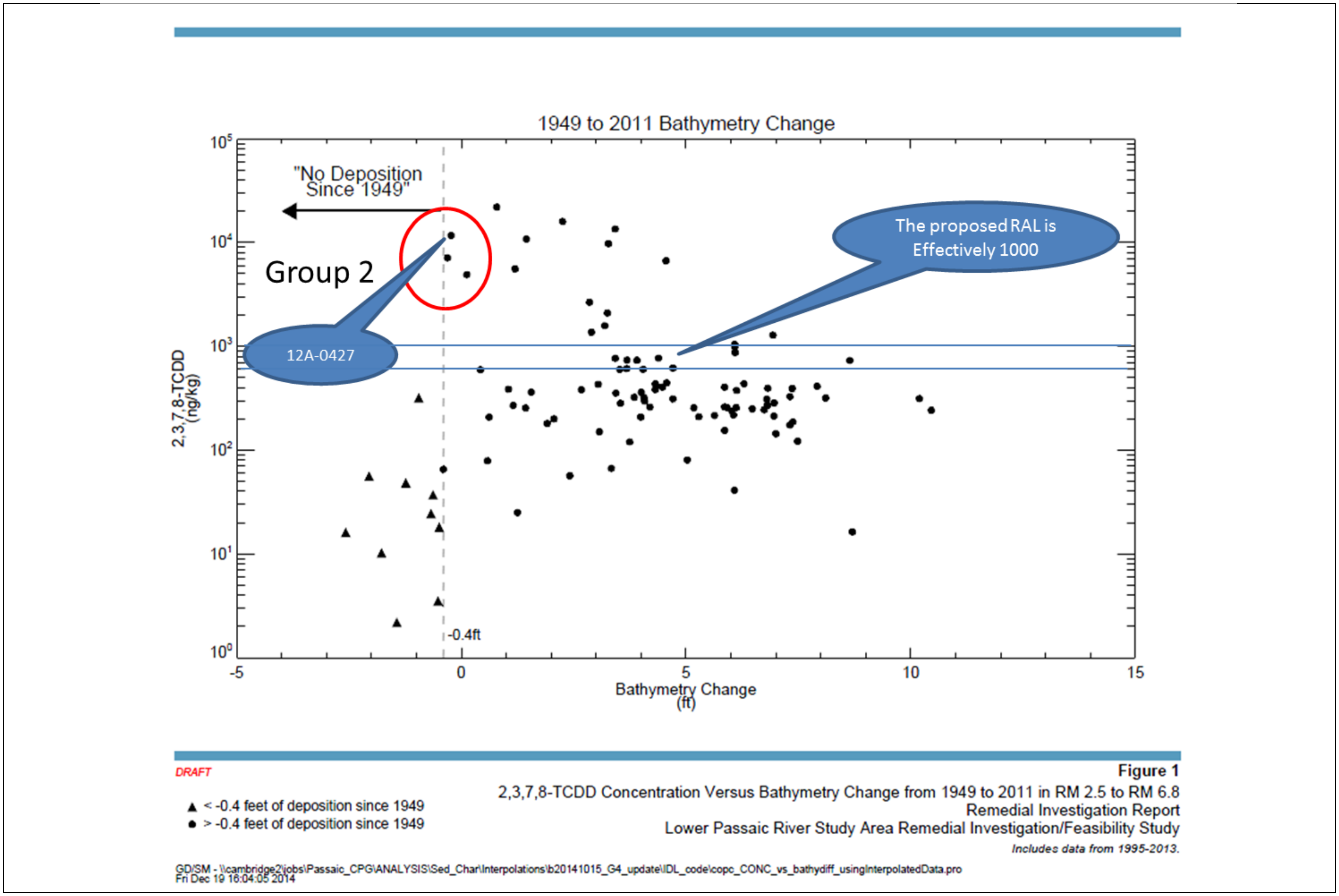


Figure 5. Influence of a small number of samples on concentration estimates within Group 3 for River Mile 2 through 6.

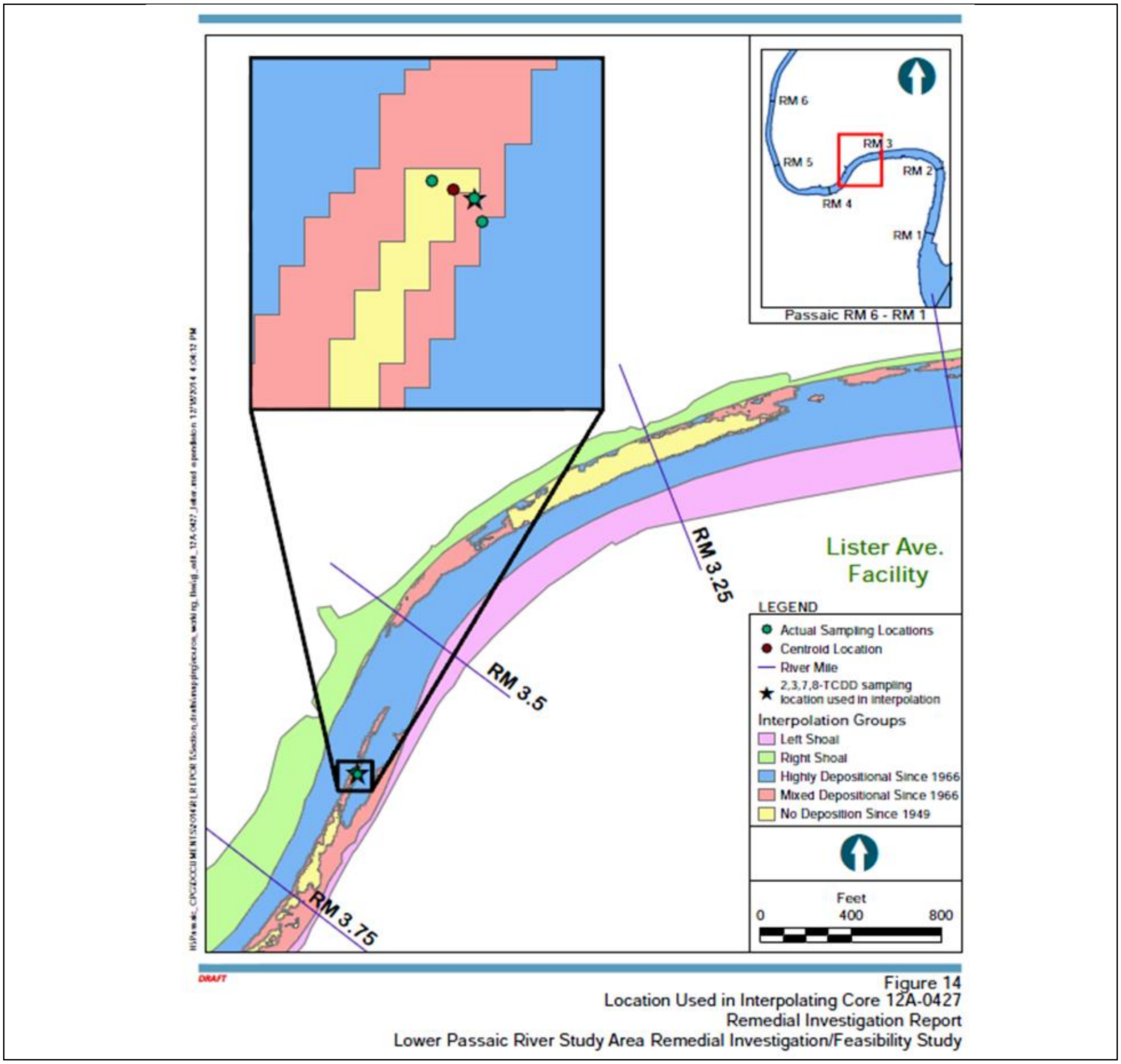


Figure 6. Disaggregation of field duplicates into individual sample values and re-assigning a single individual of the samples from Group 2 (no deposition since 1949) into Group 3 (mixed depositional). The difference in estimated erosion between these locations is approximately 3 tenths of a foot, and the distance between the centroid and revised location is less than 5 feet. The pixels in the map are 5 feet by 5 feet in size.

## 4.2 Model Robustness

The sensitivity of a predictive model to small changes in supporting data is an indicator of the robustness of the model and the likelihood that it will accurately predict values at unsampled locations. Figure 7 provides a comparison of two versions of the Mapping Approach provided to EPA in 2013 and later in 2015, both of which are based on the same data referred to by the CPG as the "2010 data".

This comparison shows that predicted values in these polygons change by over one order of magnitude, with small changes in mapping rules, even when based on the same analytical data. The left panel shows that for the 2013 map, Polygon A was predicted to have a concentration of 315 ng/kg, whereas in the newer map provided in 2015, the area is shown to have a concentrations of 104 ng/kg—a factor of 3 difference. Predictions for Polygon B increased by one order of magnitude from 1,364 ng/kg to 16,000 ng/kg (over 20 times greater than CPG's RAL). Predictions for Polygon C also changed by an order of magnitude, from 100 ng/kg to 16 ng/kg.

This illustrated sensitivity to small changes in the supporting data suggests that the Mapping Approach is unlikely to be reliable for forecasting values at unsampled locations or under future erosion and deposition. This sensitivity to new data or small changes in data handling decisions is symptomatic of models that are overly tailored to observed relationships in sample data and therefore do not generalize well, and suggests that there are likely significant inaccuracies in the estimated SWAC vs. RAL relationship.

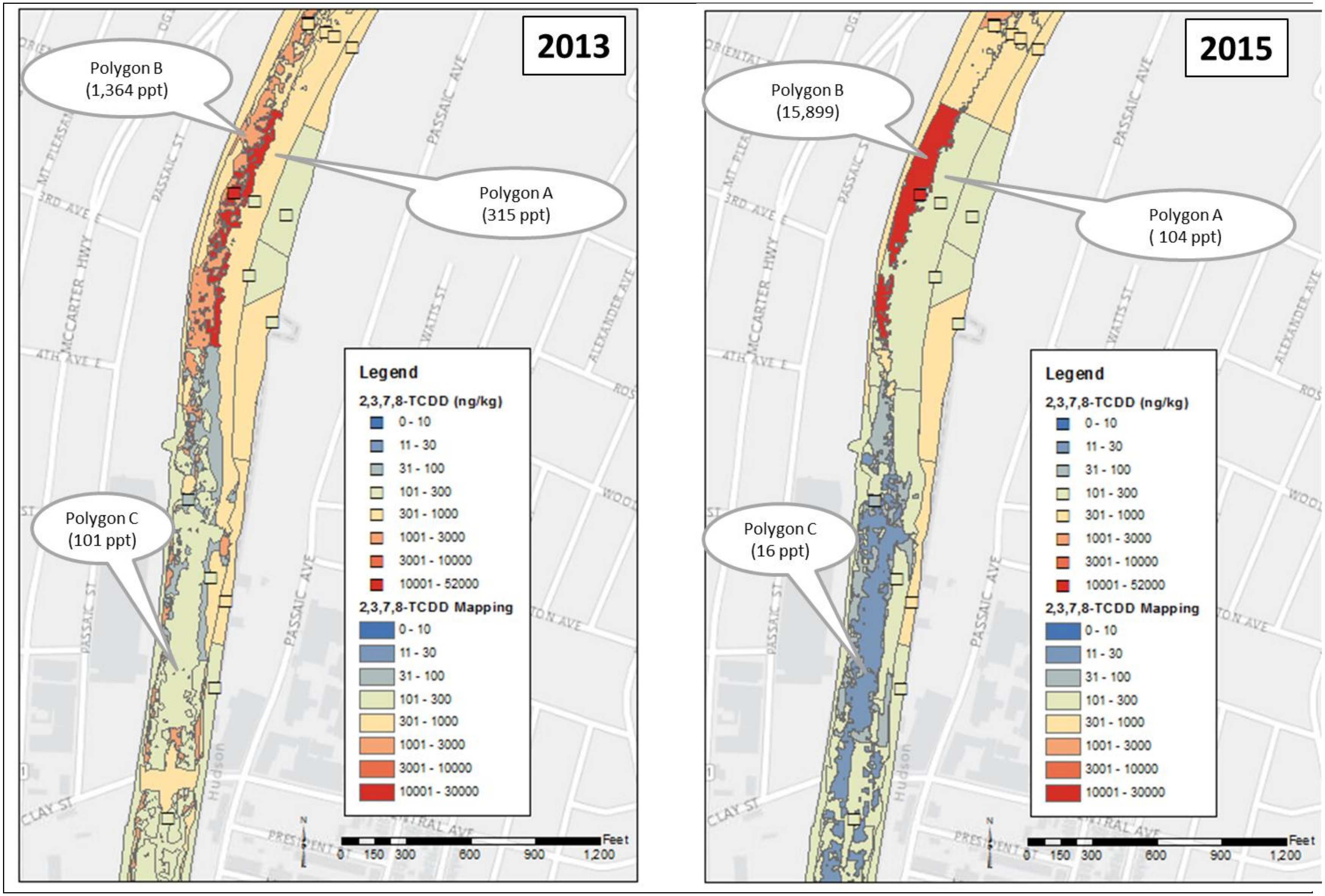


Figure 7. Mapped surfaces as presented by the CPG in 2013 (left panel) and in 2015 (right panel), both based on the 2010 dataset. Large changes in color represent sensitivity to small changes in mapping rules.

## 4.3 Additional Evaluation

The CPG's mapping approach was previously reviewed by the New Jersey Department of Environmental Protection. The results were summarized in a memo (LimnoTech, 2013) and were presented to the CPG in a meeting on September, 26 2013. A series of maps was developed illustrating concerns with the CPG procedure (LimnoTech, 2013). The LimnoTech memo concluded that *"The very sharp boundaries that AQEA draws between Group 3a, 3b and the other Group areas render the data they can use for 3a and 3b interpolation very limited and highly variable, leading to spatial patterns that defy intuition."*

Other findings in the LimnoTech memo included:

1) The groupings defined by AQEA in their 2013 version of the mapping are not well supported and may be sensitive to sharp distinctions between otherwise noisy data groups.
2) Contaminant concentration differences between erosion and deposition groups defined by the sediment transport model do not match the sharp differences between groups in the Mapping Approach.
3) Evidence that data groupings based on past erosion are reliable predictors of future erosion was not provided.
4) The small subset of data supporting development of Thiessen polygons within defined groups has limitations which lead to mapping that may be highly variable.

Figure 8, which is an updated version of Figure 2 of the LimnoTech (2013) memo, shows sample locations used to assign concentrations to Group 3 as color-filled black circles and other samples not used for assignments as solid black circles. Within any given river mile the vast majority of available samples, some within feet of Group 3 areas, are ignored, while single samples, at times thousands of feet away, are used to assign concentrations to the discontiguous polygons composing Group 3. This approach assumes sharp divisions in concentration between groups, in spite of the implicit uncertainties in the spatial boundaries of these groups and the well-documented heterogeneity of chemical concentrations in sediment.

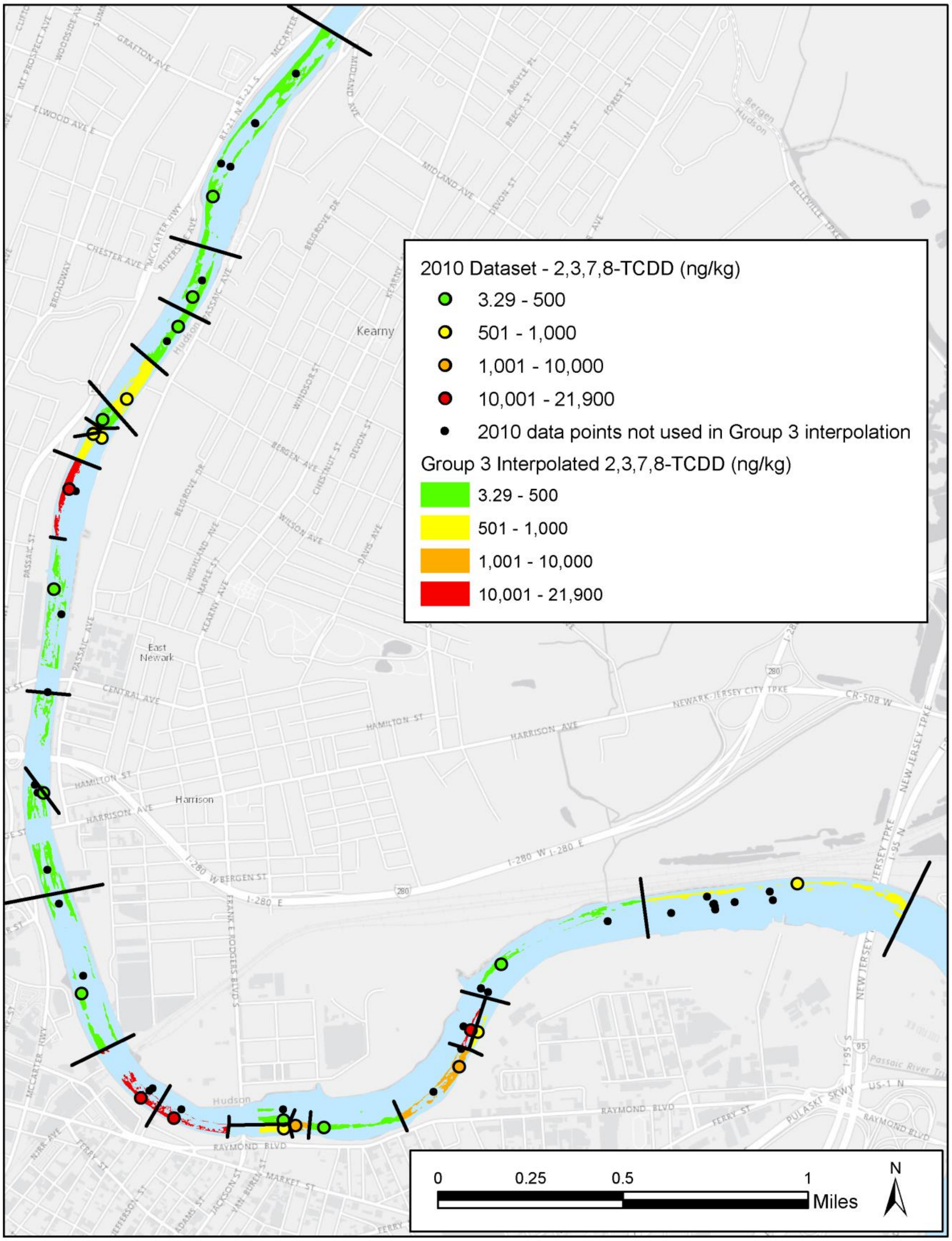


Figure 8. Updated version of Figure 2 from LimnoTech 2013 illustrating the influence of a small number of samples on concentration estimates within Group 3 in for River Mile 2 through 6. Samples indicated by black circles were excluded from estimation of Group 3 Areas in spite of close spatial proximity to interpolation areas.

# 5 Reliability of Predictions

## 5.1 Supplemental Sampling Program 2 (SSP2)

In this section, data collected in 2013 during the SSP2 sampling program to fill Data Quality Objective (DQO) 1 are compared with predictions based on the 2010 mapping. The DQOs for the SSP2 sampling program were as follows:

> DQO 1 – Provide additional characterization of the nature and extent of sediment chemistry and fill data needs above RM 8, as identified by USEPA.
>
> DQO 2 – Provide data to support system understanding, sediment surface concentration mapping, and sediment transport and CFT model parameterization.

Data collected to fulfill DQO2 would be considered biased as far as the analysis conducted herein is concerned, and thus are not included. Since the samples collected as per DQO 1 were selected to fill data gaps from previous sampling programs, they represent a relatively unbiased dataset.

Figure 9 shows a plot of predicted 2,3,7,8-TCDD concentrations on the horizontal axis and measured concentrations from the SSP2 (DQO1) sampling program on the vertical axis. The points are labeled by the geographical stratification developed by the CPG to group the data. These groups include the Left Shoal Upstream of RM 7.8 (N=15), Right Shoal Upstream of RM 7.8 (N=14), Silt Upstream of RM 7.8 (N=12) and Non-Shoal Upstream of RM 7.8 (N=16). In total there are 57 SSP2 (DQO1) sampling locations included in this analysis.

The diagonal line represents perfect prediction, and if the Mapping Approach performs well, sample and predicted values should fall on or close to the one-to-one line. In contrast, it can be seen that for any particular level of the predicted values, the range of measurements spans at least three orders of magnitude, generally from 10 ng/kg to 10,000 ng/kg. In this plot the SSP2 data are effectively uncorrelated with the CPG mapping predictions ($R^2$=0.11)[2].

Upstream of RM 8, 14 of the restricted Thiessen polygons with predicted 2,3,7,8-TCDD concentrations exceeding 500 ng/kg were subsequently sampled as part of the SSP2 (DQO1) program. Of those 14 polygons, 12 (86%) of SSP2 samples were less than the predicted value, and 10 (71%) were below the 500 ng/kg threshold (i.e. false positive predictions). This illustrates that areas identified by the Mapping Approach as exceeding the 500 ng/kg RAL actually contain lower concentration sediments.

The upper left quadrant of the plot also shows that within the 43 restricted polygons for which concentrations were predicted to be lower than 500 ng/kg, 14 (33%) exceeded the 500 ng/kg RAL (i.e. false negatives). This shows that relative to the 500 ng/kg RAL, areas predicted to be non-target areas

[2] Note that $R^2$ is shorthand notation for the coefficient of determination in regression analysis and it represents the proportion of total variance explained by the regression model. The values range between 0 and 1.0 and a value of 0.11 is indicative of virtually no relationship.

contained samples with target concentrations. Given the lack of temporal change documented previously when comparing data from 1995 with that from 2010, it can be concluded that the primary component of variation contributing to differences between the 2010 mapping and 2013 samples is spatial heterogeneity as opposed to temporal trends.

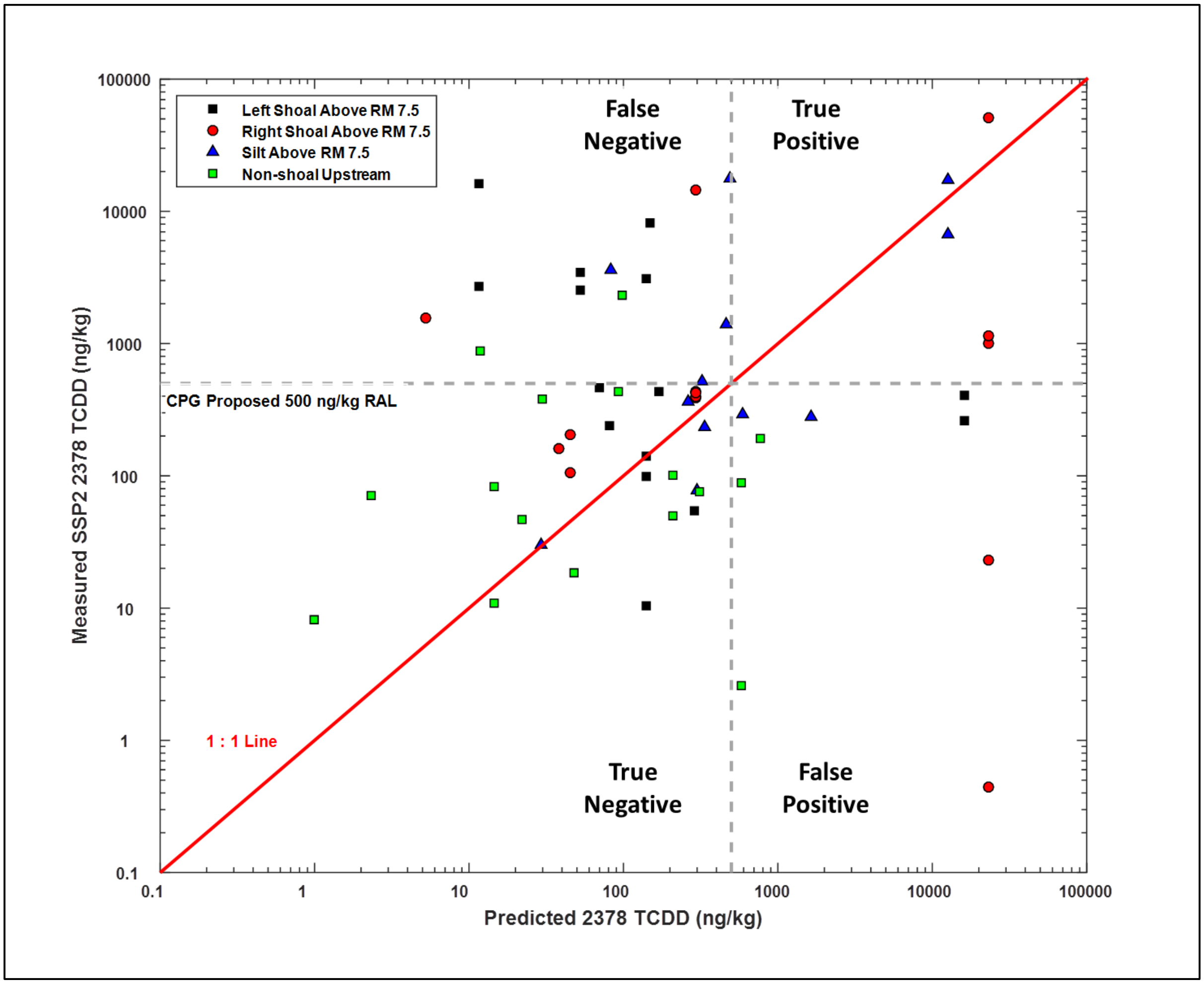


Figure 9. Measured SSP2 (DQO1) 2,3,7,8-TCDD concentration against predicted values based on constrained Thiessen polygons for 2010 data. Points right of the diagonal 1 to 1 line indicate over prediction and points left of the line indicate under prediction. The dashed gray lines represent the 500 ng/kg remedial action limit proposed by CPG. Negative and positive error labels refer to 500 ng/kg concentration.

Specific examples comparing several SSP2 samples collected in 2013 with mapped predictions based on 2010 data (AQEA, 2013) are shown in Figure 10. Samples supporting map development are depicted as squares, and SSP2 sampling locations are shown as circles. Group A shows a polygon in the Left Shoal group with a predicted value of 16,000 ng/kg but with two SSP2 samples each with concentrations less than 500 ng/kg. Group B illustrates the opposite situation where the predicted values for adjacent polygons were 12 ng/kg and 890 ng/kg, yet the SSP2 sample located on the boundary of these two

Thiessen polygons was 16,200 ng/kg, differing by more than an order of magnitude. Group D shows a polygon with a predicted value of 92 ng/kg and an SSP2 sample with 3,830 ng/kg, more than an order of magnitude higher than the predicted value.

A general pattern can be readily observed where channel areas tend to contain lower average concentrations than the shoals, but within these general groupings there is significant variation (up to 4 orders of magnitude in range) in concentrations above and below the proposed 500 ng/kg RAL.

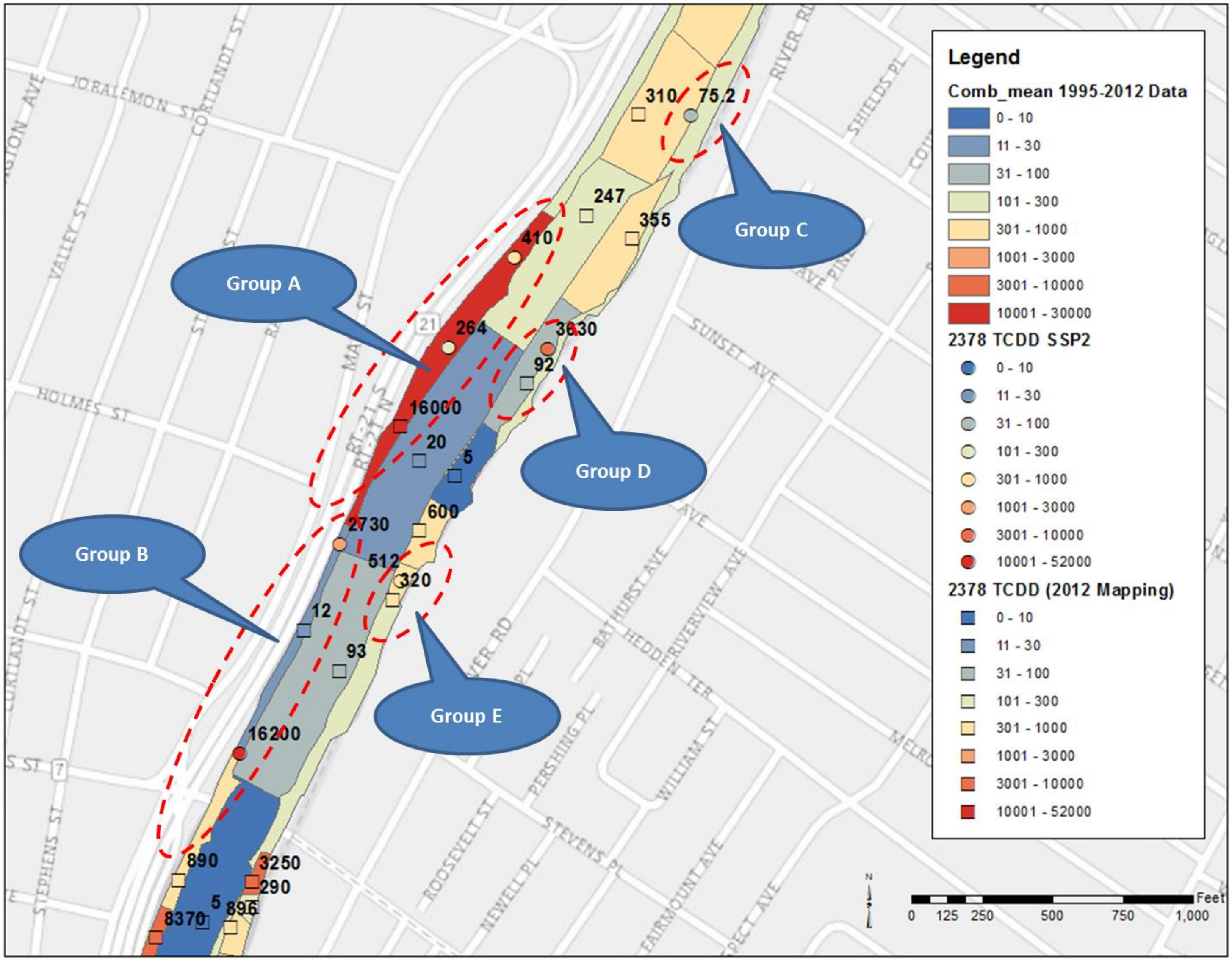


Figure 10. Comparison of SSP 2 samples with mapping based on 1995 through 2012 data.

## 5.2 Group Mean Contrasts

As described in more detail in Section 3, the primary groups identified for the Mapping Approach are left shoals, right shoals, Groups 2, 3, and 4 downstream of RM 7.8, and non-shoal and silt groups upstream of RM 7.8. However, statistical comparisons of contaminant concentrations among groups were not provided. In this section, the geometric means of contaminant concentrations are compared among the groups, based on 2010 SSP data and separately based on SSP2 (DQO1) data. Differing concentrations among groups would support the concept that the Mapping Approach may be useful for prediction and RI/FS evaluations.

In addition to the primary grouping, more localized subgroups ordered from downstream to upstream were developed. For example, one such subgroup is labeled “SI_B” representing silt area B and is coincident with the contaminant deposit at RM 10.9 which was the subject of a removal action. In total there are 26 such groups containing surface sediment samples. These subgroups range in size from approximately 2 acres to over 200 acres and average 37 acres. The sampling density is approximately 1 sample per 2.8 acres. The groups break the data into subsets by left shoal, right shoal, navigation channel and silt approximately by river mile upstream of RM 7.8 and based on erosional groups downstream of RM 7.8.

Estimated geometric mean concentrations in Figure 11 segregated by the defined subgroups and organized by sampling program SSP (left panel) and SSP2 (DQO1; right panel) show that the SSP sample non-shoal areas upstream of RM 8 generally have lower concentrations than shoal areas upstream of RM 8, and that non-shoal areas upstream of RM 8 also have lower concentrations than all areas in the lower 8 miles, with the exception of Group 2.

Generally speaking, few areas exhibit the large ratios of means that would be required to achieve large SWAC reductions with the small remedial footprint proposed. With the exception of areas upstream of about RM 13 (LS_J, NS_D, NS_E and NS_F), most group geometric means are within one order of magnitude or less, as opposed to the ratio of nearly 26 to one necessary to support a correspondence between a RAL of 500 ng/kg and a post-remedial SWAC of 150 ng/kg with a remedial footprint of 15% of the area. Analyses used to support this correspondence are based on much smaller groups of data (essentially individual samples), which as discussed above may produce unreliable results.

Geometric mean 2,3,7,8 TCDD concentrations from the SSP2 (DQO1) program, all of which were upstream of RM 7.8, were higher than geometric means based on the SSP program for groups classified as Silt, Non-shoal and Left Shoal. Geometric mean 2,3,7,8 TCDD concentrations from the SSP2 data from the Right Shoal were comparable with those based on the SSP collection. Estimated differences between Non-shoal and Shoal areas based on SSP2 data, which did not inform group definition, were smaller than forecast by the SSP data that were used to define the groups. This illustrates the issue pointed out by Smith (1983) wherein use of the dependent variable in defining group membership leads to inaccuracies in subsequent estimation and forecast

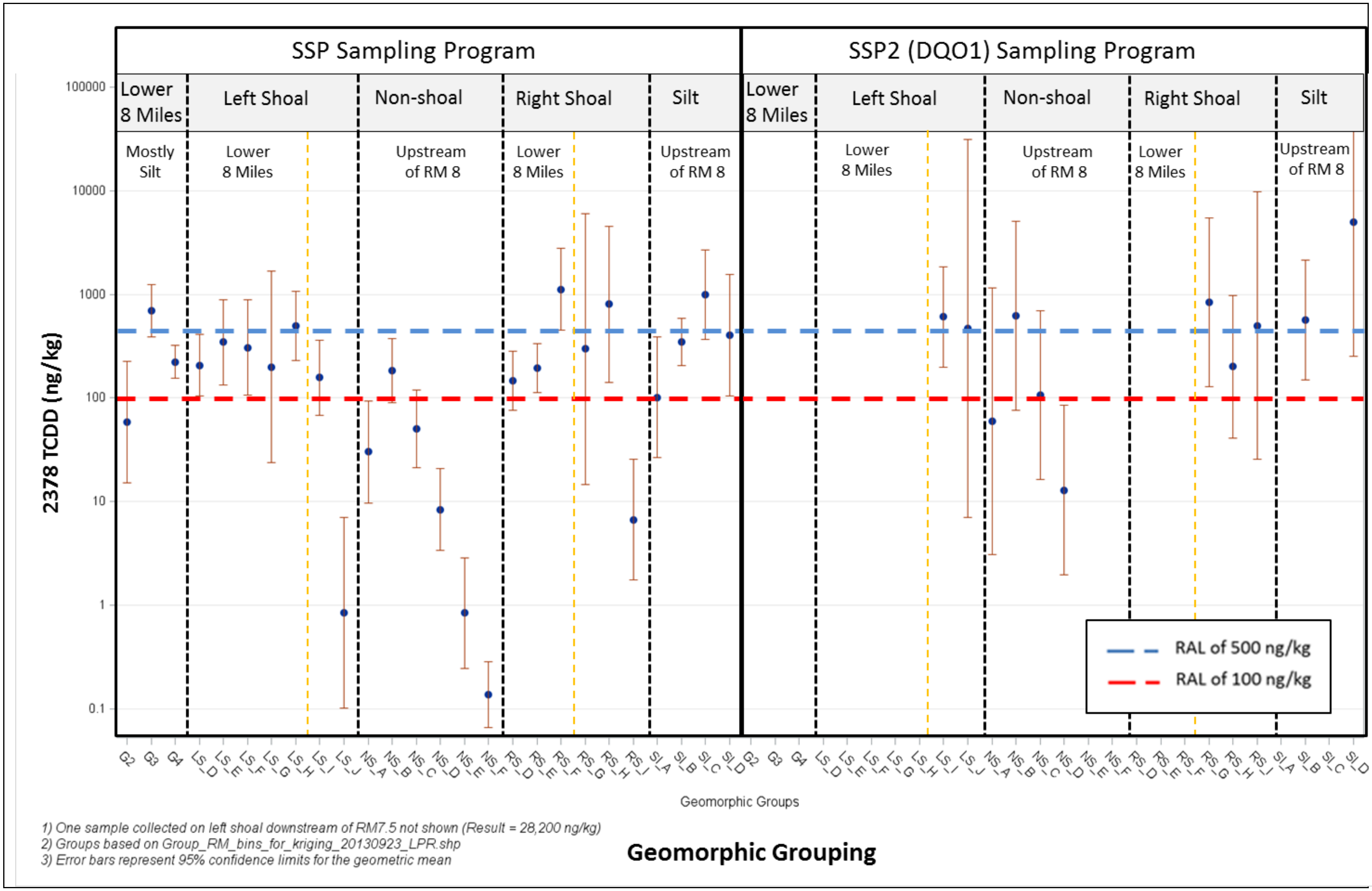

Figure 11. Geometric mean 2,3,7,8-TCDD by geomorphic and erosional groups defined by the CPG.

# 6 Evaluation of Post-Remedial SWAC Forecast

Calculations used to support the proposed 500 ng/kg RAL and 150 ng/kg PRG do not account for the fact that the calculations are based on individual point samples representative of a few square inches of surface area, while remediation occurs at a much larger scale (on the order of acres). This problem of using point scale data to make inferences to larger acre scale decision units is commonly referred to in the Geostatistical literature as the "*Change of Support"* problem (Journel and Huijbregts, pg 15, 1978).

This problem has been thoroughly discussed in relation to mine planning, with respect to making the faulty assumption that all the gold exceeding certain economic-driven grades could be removed, while leaving behind all of the material that is below cutoff-grade, using large-scale equipment. In the context of the LPRSA, this error is likely to result in overstatement of the effectiveness of remedial options and understatement of the surface area and volume of sediment that needs to be addressed.

The true effect of a remedial action on surface concentrations is a function of the actual average concentration in the areas identified for removal. The Mapping Approach focuses on the development of individual Thiessen polygons, each representing an interpolated average concentration based on a single sample per polygon. This leads to the "change of support" problem described above. While the use of interpolation to represent average concentrations based on limited data is a common practice and is often valid, the limitations and potential pitfalls of this approach need to be understood and, as possible, accounted for.

This section evaluates the limitations of this approach as utilized for the RI/FS by:

1) illustrating the procedure used through example;
2) testing the accuracy of the forecast of post-remedial SWAC, also by example; and
3) conducting a small simulation to verify that the findings illustrated by examples in evaluation methods 1 and 2 are the rule as opposed to the exception when forecasting SWAC vs RAL relationships.

## 6.1 CPG Method Description

The Mapping Approach is applied to forecasting the post-remedial SWAC by simulating the effects of a remedy. This is done by iteratively removing polygons with the highest concentrations and re-averaging hypothesized post-remedial concentrations. The CPG's procedure (the Forecasting Procedure) follows 5 steps as follows:

1) Replace the highest concentration value with a post-remedial concentration (approximately zero ng/kg).
2) Recalculate the SWAC after replacement of the highest value.
3) Check to see if the resultant forecast SWAC is below risk-based thresholds, or if the decline in SWAC is beyond the point of diminishing returns.
4) Calculate associated remedial footprint and volume of contaminated sediment to be removed.
5) Repeat Steps 1-4 until risk management thresholds are reached, or a point of diminishing returns is identified.

In order to confirm the understanding of the Forecasting Procedure, EPA followed the outline above and found a similar correspondence between a 500 ng/kg RAL and a 145 ng/kg post-remedial SWAC. In the following sections the reliability of this procedure is evaluated.

## 6.2 Simulated Example using data from River Mile 10.9

A deposit along the right bank of the LPRSA at RM 10.9 has been studied in detail with a series of sampling programs, starting in 2008/2009 with a coarse sampling of the area, followed by subsequent rounds of sampling. The sampling density for the 2008/2009 events was approximately 2 to 3 acres per sample which is similar to existing RI data throughout the rest of the LPRSA. Samples collected in 2008 and 2009 originate from the 2007 EPA EMBM (N=4), 2008 CPG Low-resolution Coring Program (N=7), and the 2009 phase of the FSP2 Benthic Sediment Sampling (N=15). For purposes of this example, we refer to these data from 2008 and 2009 as the "RI data". These data, combined with data from subsequent investigations with generally higher sampling density (referred to herein as the pre-design data), provide an opportunity to explore the accuracy of the SWAC vs RAL calculations forming the basis of the CPG's remedial alternatives analysis. Implementing the Forecasting Procedure based only on the RI data described above is representative of the estimates the CPG has developed throughout the LPRSA, whereas repeating the calculations with all of the data spanning 2008 to 2013 provides a more accurate estimate of the actual reduction in SWAC.

EPA used extensive data from RM 10.9 to evaluate the performance of the Forecasting Procedure, in a situation where data are adequate to support the evaluation. The results illustrate how the Forecasting Procedure functions when used to forecast remedial performance with sampling densities of approximately one sample per 2 to 3 acres. This illustration does not constitute an evaluation of the performance of the RM 10.9 removal action, and is not meant to suggest that the objectives of the removal action were not met. This is being used as an example because it provides a relatively large dataset collected for two distinct purposes – one to determine the general nature and extent of contamination for the RI and the other to design a removal action.

The simulation focused on a stretch of the river running from approximately RM 10.5 to RM 11.7, comprised of an area of 59.7 acres referred to as the Focus Area (Figure 12). This Focus Area contains 116 sediment sampling locations with samples collected at the surface through 10 sampling programs undertaken from 2008 to 2013 (Table 1). The 116 locations in the Focus Area are identified with 4 depositional groups: Left Shoal, Non-Shoal, Silt, and Right Shoal, delineated by the CPG based on their interpretation of bathymetric gradients, patterns in sediment texture and analytical chemical results.

| Table 1. Sediment investigation and sample counts. | | |
|---|---|---|
| **Study Phase** | **Year** | **No. of Samples** |
| 2007 EPA EMBM | 2008 | 4 |
| 2008 CPG Low-resolution Coring Program | 2008 | 7 |
| 2009 FSP2 Benthic Sediment Sampling | 2009 | 15 |
| 2010 FSP2 Benthic Sediment Sampling | 2010 | 1 |
| 2011 River Mile 10.9 Field Investigation | 2011 | 54 |
| 2012 Focused Sediment Investigation | 2012 | 3 |
| 2012 Focused Sediment Investigation JDG Splits | 2012 | 2 |
| 2012 River Mile 10.9 Addendum A | 2012 | 15 |
| CPG Second Supplemental Sampling Program (SSP2) | 2013 | 13 |
| CPG Supplemental Sampling Program | 2012 | 2 |

### 6.2.1 Post-Remedial SWAC Forecast Based on 26 Samples

The initial step of the analysis used 26 samples collected from 2008 to 2009, i.e. the RI data. Following the CPG's contaminant mapping approach described above, Thiessen polygons were constructed within each of the four depositional groups based on the 26 sample locations (Figure 12).

The calculations show that using the RI data alone, the Focus Area has an estimated SWAC of 778 ng/kg with 19.6 acres exceeding the RAL. Applying the CPG's procedure for forecasting the effect of remediation based on these RI data, remediating the polygons (each containing a single sample) identified as exceeding the RAL would result in a reduction of the estimated SWAC to 96 ng/kg (Table 2).

| Table 2. SWAC Forecast Based on 26 Thiessen Polygons | |
|---|---|
| | **RI Data (2008 to 2009)** |
| Estimate of Pre-cleanup SWAC (ng/kg) | 788 |
| Estimate of Post-cleanup SWAC (ng/kg) | 96 |
| Estimate of Area Exceeding RAL (acres) | 19.6 |
| Estimated SWAC of Area Exceeding RAL (ng/kg) | 2104 |

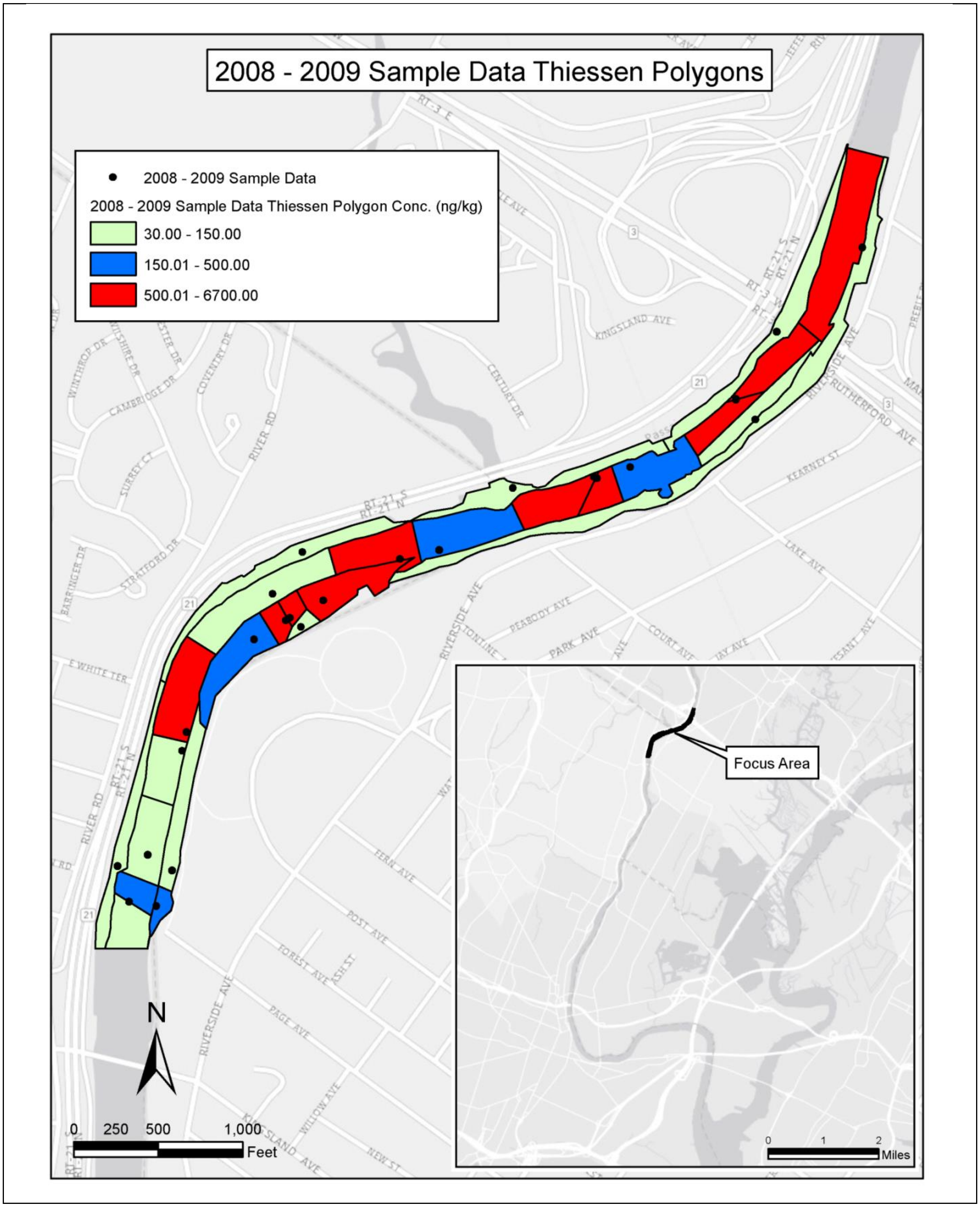


Figure 12. Thiessen polygons for RI surface samples restricted to CPG mapping groups near River Mile 10.9 (The Focus Area).

### 6.2.2 Accuracy Evaluation Based on 116 Samples

The second part of the analysis was to compare the post-remedial SWAC forecast based on the RI data to revised estimates using all data, including the pre-design data (116 samples total). For this comparison, EPA intersected the Thiessen polygons developed from the RI data with the full data Thiessen polygons and calculated several numbers including 1) a post-remedial SWAC, 2) the total area with concentrations lower than the RAL that was removed, 3) the total area remaining with concentrations above the RAL, and 4) the SWAC of the area that was removed. These calculations were performed based on the concentrations from the full sample data Thiessen polygons clipped to the Thiessen polygons determined by the RI sample data. In other words, EPA estimated the "actual" outcome of remedial action based on the RI data, using a more accurate estimate of the concentrations in each Thiessen polygon provided by the full set of 116 samples (Figure 13, Table 3).

**Table 3. Comparison of forecast and more accurate SWAC estimates at RM 10.9.**

| Estimated Metrics | Forecast Based on RI Data Polygons and Data | Estimate Based on RI Data Polygons and All Data |
|---|---|---|
| Pre-Remedial SWAC (ng/kg) | 788 | 1,410 |
| Post-remedial SWAC (ng/kg) | 96 | 805 |
| Percentage SWAC Reduction (%) | 88% | 43% |
| SWAC Outside Remedial Footprint (ng/kg) | 142 | 1,199 |
| SWAC Within Area Exceeding RAL (ng/kg) | 2,104 | 1,839 |
| Remediated Area Exceeding RAL (acres) | 19.6 | 10.2 |
| Unremediated Area Below RAL (acres) | 40.0 | 35.5 |
| Unremediated Area Exceeding RAL (acres) | 0 | 4.5 |
| Remediated Area Below RAL (acres) | 0 | 9.5 |

Figure 13 shows that the assumption that sediments within the remedial footprint exceed the RAL, and that those outside the footprint are lower than the RAL, does not hold true in practice. Some sample data exceeding the RAL (red circles) are outside the remedial footprint identified by blue Thiessen polygons, and other samples within the remedial footprint (red Thiessen polygons) have concentrations below the RAL. These errors in delineation are not unexpected, but an evaluation of the potential biasing effects of these incorrectly delineated sediments on forecasts of post-remedial SWAC has not been provided.

This example shows that the benefit of remediation in the Focus Area is overstated by the Forecasting Procedure. Based on the RI data only, the SWAC is forecast to be reduced by 88%, nearly a factor of 10. However, when the more complete data set is used to estimate the remedial benefit of the same remedial footprint (i.e. based on RI polygons) the more accurate estimate of SWAC reduction is just 43%, which is less than a factor of 2.

Table 3 shows a comparison of the forecast of post-remedial SWAC based on the CPG calculation procedure, compared with the more accurate estimates based on the full data set applied to the Thiessen polygons based on RI data. This comparison shows that for this specific area near RM 10.9, the actual post-remediation SWAC based on the RI Thiessen polygons would be 805 ng/kg, a SWAC more than eight times greater than forecast. Of the 19.6 acres identified for remediation, 10.16 acres (52%) actually exceeded the RAL whereas the remaining 9.5 acres (48%) had concentrations below the RAL.

Figure 14 shows the spatial distribution of polygons that were or were not targeted for remediation and their 2,3,7,8-TCDD concentrations based on the full set of 116 samples. According to this evaluation, blue polygons were correctly identified for remediation, and green polygons were correctly identified as not needing remediation, based on subsequent rounds of data collection. Red polygons indicate samples that exceeded the 500 ng/kg RAL but were not targeted for remediation, whereas tan polygons indicate samples that were below the 500 ng/kg RAL but were incorrectly targeted for remediation. In non-target areas, an estimated 4.5 acres exceeded the RAL, representing 44% of the 10.16 acres of target material exceeding the RAL.

SWAC was underestimated in areas not targeted for removal due to red areas in Figure 14 that should have been targeted for removal, but were not. These areas are primarily along the left and right shoals but notably within the downstream half of the River Mile 10.9 deposit. The estimated SWAC in the non-remedial areas based on RI data was 142 ng/kg in contrast to the more accurate estimate based on the full data set of 1,199 ng/kg—nearly a factor of 10 higher than predicted using the Forecasting Procedure. Notably, this value also exceeded the pre-removal SWAC estimate based solely on RI data.

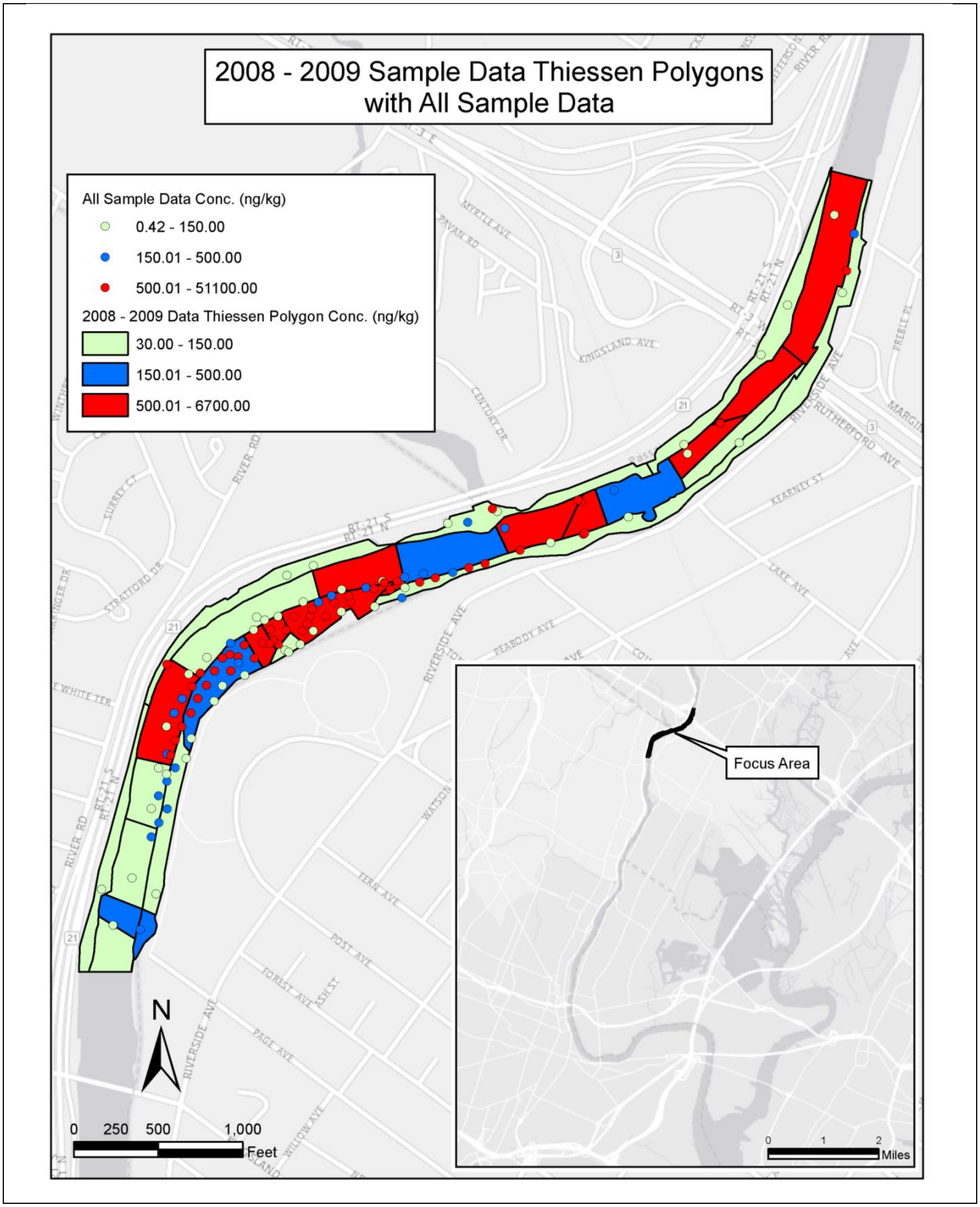


Figure 13. All samples collected from 2008 through 2013 overlaid on RI data Thiessen polygons showing samples exceeding the 500ng/kg RAL outside the identified target areas and areas below the RAL within the target areas.

The SWAC was overestimated in the removal areas because the Forecasting Procedure did not account for sediments within the footprint having concentrations below the RAL. These are indicated by tan polygons in Figure 14 with their area being predominantly associated with non-shoal areas, but also with a small number of samples within the RM 10.9 deposit. SWAC in the removal area was forecast based on the RI data to be 2,104 ng/kg, but after accounting for these below-RAL samples within the footprint, the more accurate estimate based on all the data was 1,839 ng/kg.

The preceding illustration is based on a single example along the LPR where RI sampling density was approximately 1 sample per 2 to 3 acres, which is similar to the sampling density over most of the LPRSA. Figure 15 shows the areas to the south and north of the Focus Area with all samples collected to date, 2008-2013 showing qualitatively similar sample density. This suggests that the resolution of current data supporting the CPG's forecast of post-remedial SWAC based on a range of RALs is likely to be subject to inaccuracies on the order of those illustrated by example at River Mile 10.9.

In the next sections, a small simulation study is conducted, further evaluating the CPG's Forecasting Procedure, testing the degree to which RM 10.9 is a representative example or simply an anecdotal situation.

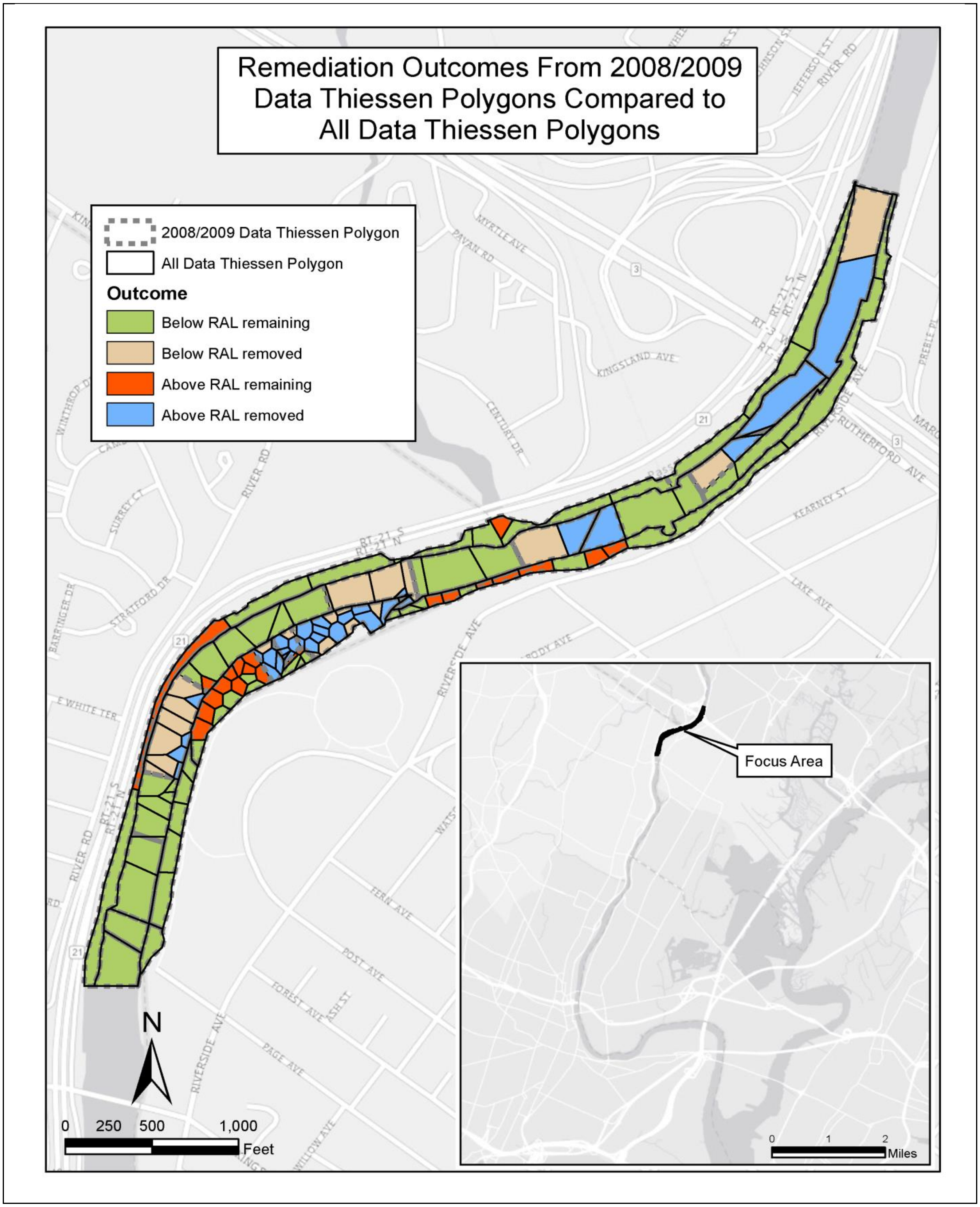


Figure 14. Polygon disposition based on 2008 to 2009 data remedial footprint. Orange polygons represent untargeted sediments exceeding the RAL and tan areas represent concentrations below RAL that were targeted for removal.

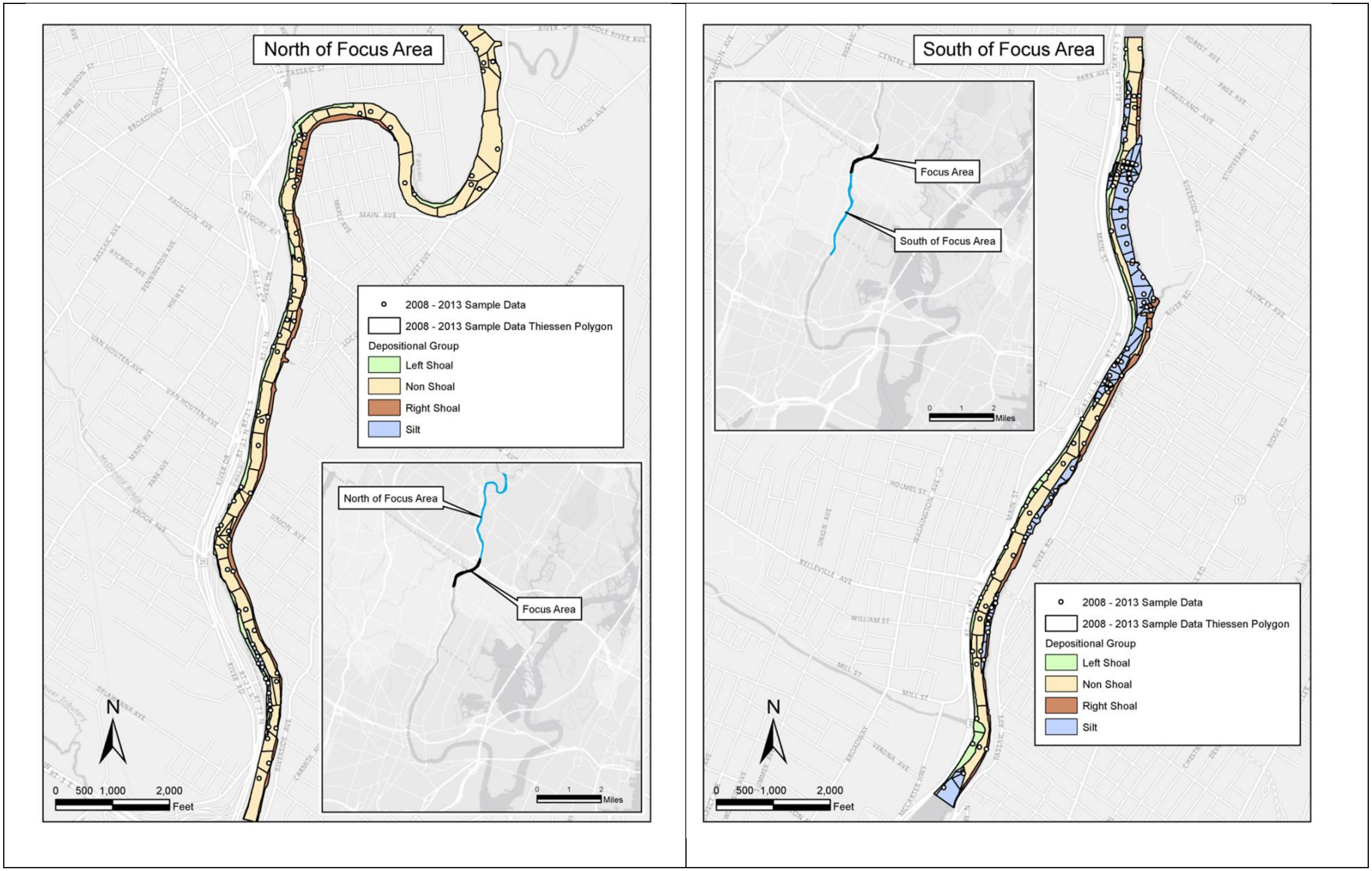


Figure 15. Samples in River Mile 12-15.6 (left panel) and River Mile 8-10.5 (right panel) showing the sampling density to be approximately 20 samples per mile (72/(15.65-12)).

## 6.3 Simulation Study Generalizing RM 10.9 Illustration

To evaluate whether the RM 10.9 example is overly anecdotal and not reflective of a general problem, additional lines of evidence were developed based on computer simulation studies that allow many thousands of replicate delineation problems to be evaluated to find general patterns in the performance of the SWAC forecasting problem, including:

1) Is the bias noted at RM 10.9 pervasive or anecdotal?
2) Is post-remedial SWAC systematically understated, or are positive and negative errors equally likely?
3) Do spatial correlations mitigate or exacerbate biases?
4) Does increased sampling density mitigate biases?
5) How can potential sampling and analysis biases be factored into the FS process to equilibrate comparisons of disparate remedial options?
6) Are there other alternative approaches that may provide understanding of the bias present in any given sampling program, and can the bias be corrected?

In this section a computer simulation study is used to test the Forecasting Procedure under controlled conditions, wherein actual pre- and post-remedial SWAC values are known. Simulation uses computer intensive methods to evaluate performance of statistical procedures against a known true condition. The basic idea of a computer simulation is to generate many (thousands) of data sets, statistically similar to actual data, followed by application of the calculation procedure to each set. Because the synthetic data are exhaustive, the parameters of interest (i.e. SWAC) are fully known so that the accuracy of estimates can be established. Simulation is a standard tool used by statisticians when performance evaluations are mathematically intractable and cannot be carried out with real sample data alone (Burton et al. 2006). Statistical simulation methods have been used by EPA to develop recommendations for estimating confidence limits (USEPA, 2013) and have also been used at other Superfund sites (Kern et al. 2009).

### 6.3.1 Simulation Algorithm

A large number of synthetic maps were simulated, each including three zones with contamination levels similar to those observed in the left shoal, right shoal and channel of the LPRSA in the vicinity of RM 9 to RM 11. The distribution has lower concentrations in the channel and higher concentrations on the left and right shoals, consistent with the situation found at the LPRSA.

The simulation approach follows 8 steps:

1) Simulate mean zero spatially correlated residual process
2) Add the zone means to each of three separate zones (Shoals and channel) and exponentiate
3) Subdivide each strip into 8 rectangular decision units (24 total units)
4) Select one sample from each of the 24 units
5) Apply CPG SWAC vs RAL Forecasting Procedure (Described in Section 6.1)
6) Calculate actual SWAC as mean of Decision Unit (DU) synthetic surface after setting to zero
7) Calculate ratio (R) of Actual to Forecast SWAC
8) Repeat 1000 times and summarize results graphically

Synthetic distributions within each zone were lognormally distributed with cumulative distributions for each of the three zones as shown in Figure 16. The simulation area was subdivided into 24 decision units, each 64 feet long and 20 feet wide, with boundaries restricted to channels or shoals (Figure 17).

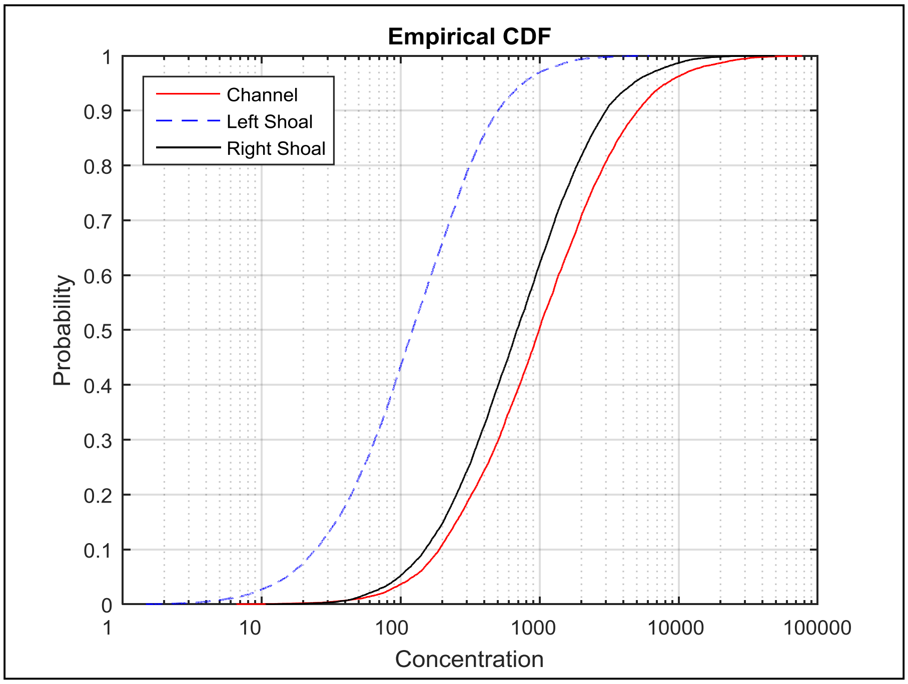


Figure 16. Cumulative distribution of synthetic 2,3,7,8-TCDD concentrations with average concentrations of 2,297 ng/kg, 228 ng/kg and 1,367 ng/kg in the left shoal, channel and right shoal respectively.

Figure 17 provides a plan view of one simulated map with concentrations given by the cumulative distributions from Figure 16 and with range of influence of 64 feet in the long flow direction and 12.8 feet in the cross flow direction. The long- to cross-flow ratio of anisotropy is 5 to 1. The nugget effect was assumed to be zero. These simulation parameter settings are similar to those observed at the LPRSA, in that each decision unit is approximately one range of influence in length.

The true synthetic average is posted on each decision unit, and the value of a single sample within the unit is posted on the diagram in parentheses. In Figure 17, the true average value, for this particular synthetic map, is colored red (N=8) for decision units for which the single sample value exceeds the 500 ng/kg RAL, indicating units targeted for removal.

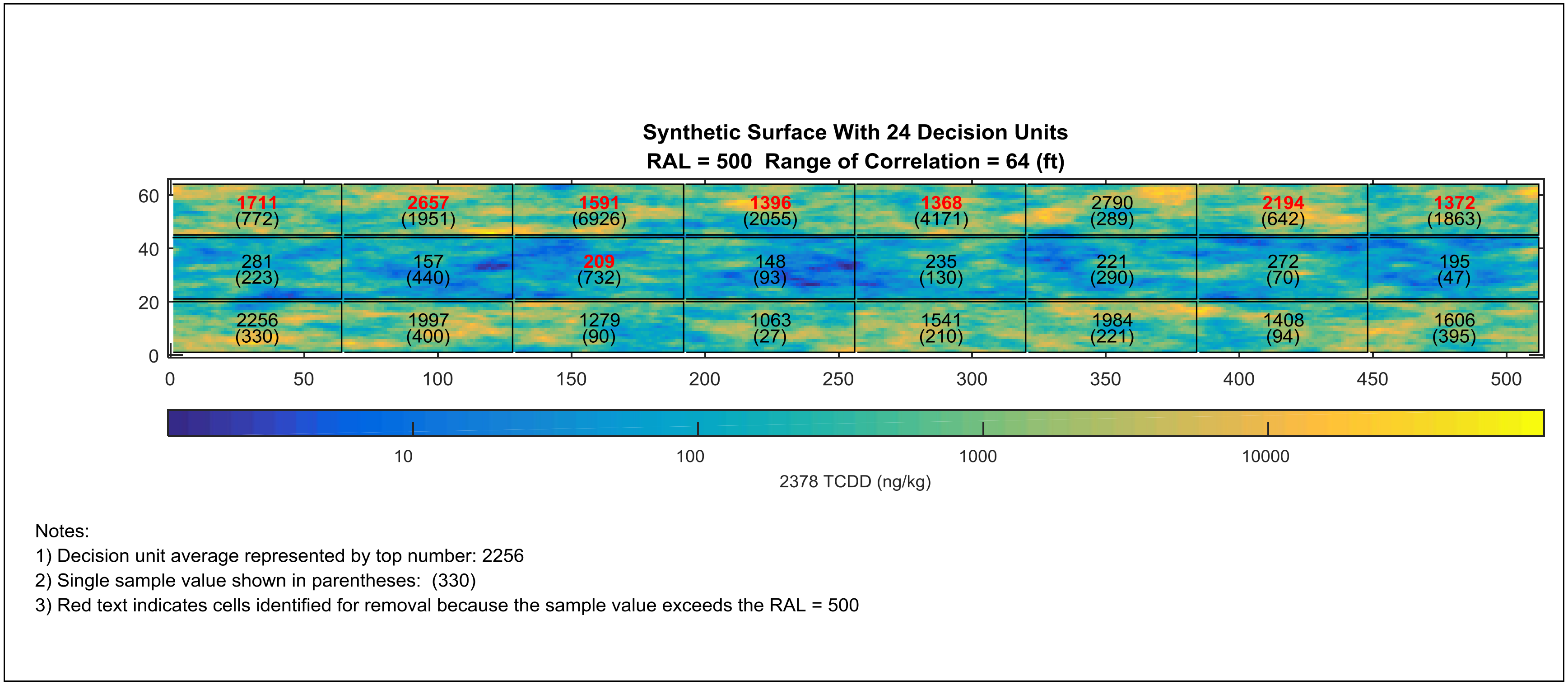


Figure 17. Synthetic 2,3,7,8-TCDD concentrations with average concentrations of 2,297 ng/kg, 228 ng/kg and 1367 ng/kg in the left shoal, channel and right shoal respectively. The range of correlation is 64 feet in the long flow direction—equal to the distance between decision unit centroids. The true Decision Unit mean is listed within each decision unit.

### 6.3.2 Actual vs Predicted Decision Unit Averages

In the majority of DUs (or equivalent Thiessen polygons) the forecast overstated the average in target areas, and understated the average in non-target areas. Taken together, the post-remedial SWAC was understated (i.e. overstating the apparent remedial benefit) in this synthetic example. Figures 17 and 18 provide a summary of the results from a single simulated map. In this instance, the actual DU average is lower than forecast (i.e. single sample) in 5 of 8 units targeted for removal and the average was higher than forecast in 13 of 16 units not targeted for removal. Each true average is plotted against the forecast for this particular example in Figure 18. The actual means tend to be higher than predicted values in the non-target group (above the 1 to 1 line). Conversely, the actual means in the target group tend to be lower (below the 1 to 1 line) than the forecast values in the target groups. This result is consistent with the result illustrated at RM 10.9. The following section discusses the simulation of many such synthetic maps to derive the general statistical properties of bias and precision for the CPG's procedure.

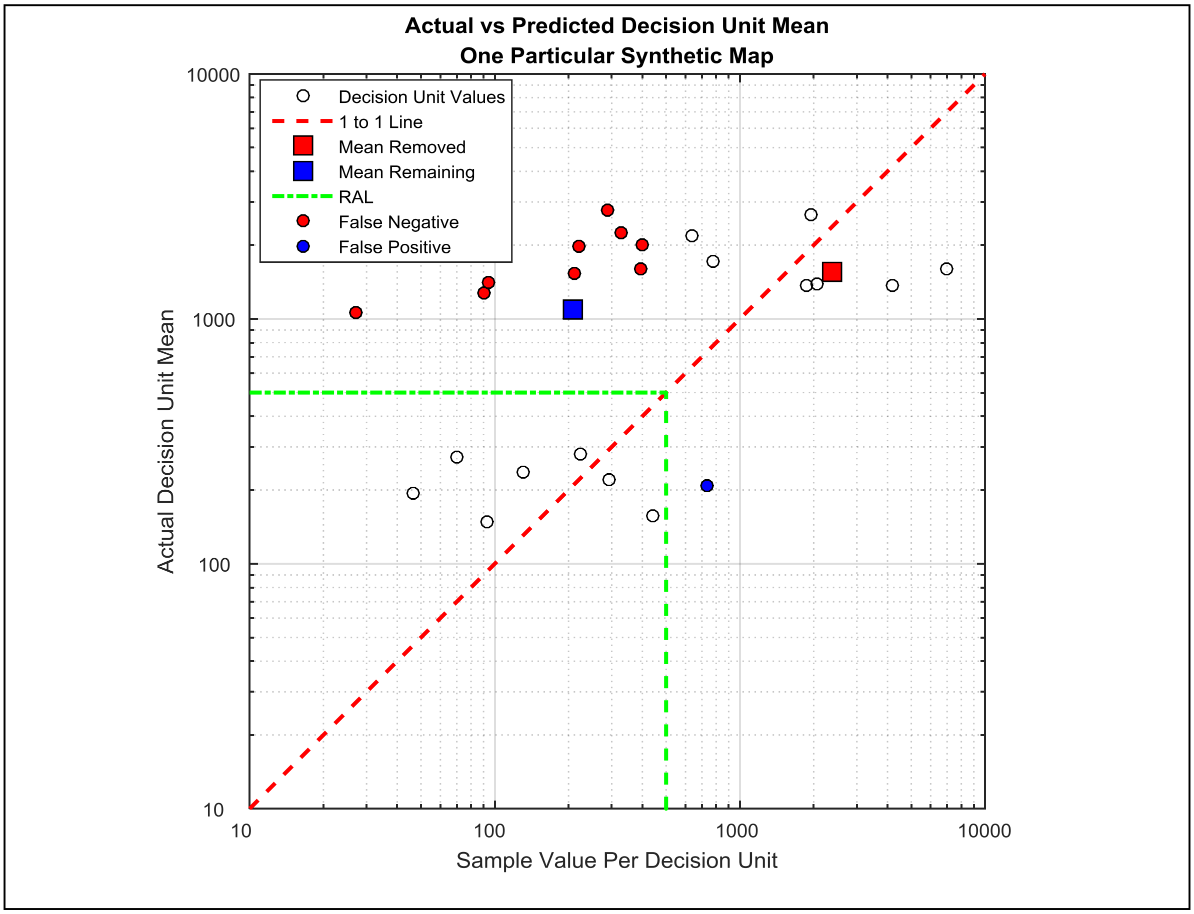


Figure 18. Average synthetic 2,3,7,8-TCDD concentrations vs single sample value determining removal decisions. The blue square represents the true average in the non-removal areas plotted against the forecast concentration based on single sample per DU. The red square represents the true average plotted against the forecast average based on one sample per DU. Filled red circles represent DUs that were falsely identified as non-target areas and filled blue circles represent DUs that were falsely identified as exceeding the RAL.

### 6.3.3 Properties of Post-Remedial SWAC Forecast

Results from simulating the CPG's SWAC forecasting method are summarized in this section. The upper left panel of Figure 19 shows that the actual post-remedial SWAC exceeded the forecast in 999 out of 1000 synthetic maps. Similarly, the actual remedial footprint was greater than forecast for the clear majority of the synthetic maps shown in the upper right panel of Figure 19.

The lower panels of Figure 19 show the statistical distribution of the ratio of the actual to forecast post-remedial SWAC (lower left panel) and actual to forecast footprint size (lower right panel). The ratio of true to forecast footprint size ranged from 1 to 11 and averaged 4.6. The ratio of true to forecast SWAC ranged from 1 to 3 and averaged 1.5.

The results from this simulation study suggest that the bias observed for the specific RM 10.9 example is generally due to the statistical biases in the calculation procedure used as opposed to being anecdotal to RM 10.9. These results suggest that CPG's Forecasting Procedure understates the post-remedial SWAC and the footprint size associated with a 500 ng/kg RAL. This indicates that the correspondence between a 500 ng/kg RAL and a 150 ng/kg PRG is likely to be inaccurate and cannot be relied upon to support decision making for the LPRSA. The CPG's SWAC Forecasting Procedure likely overstates the effectiveness of a remedial action based on a RAL of 500 ng/kg.

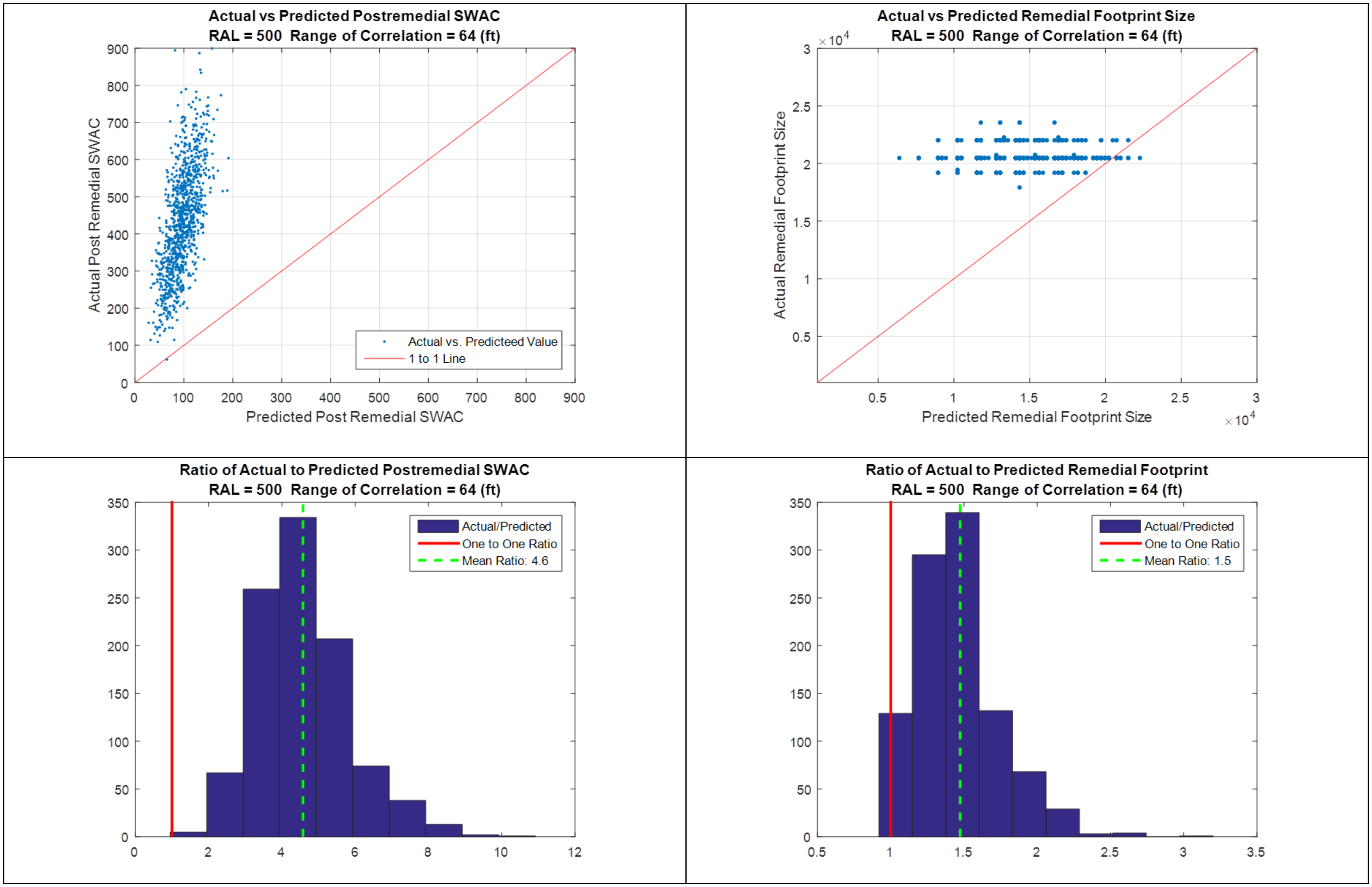


Figure 19. Actual vs predicted post-remedial SWAC (upper left panel ) and remedial footprint size (upper right panel) and ratio of actual to predicted post-remedial SWAC (lower left panel) and remedial footprint size (lower right panel) for 1000 synthetic 2,3,7,8-TCDD realizations subjected to the hill-topping algorithm used to develop the 500 ng/kg RAL proposed by the CPG.

# 7 Conclusions

The evaluations described herein show that the Mapping Approach is sensitive to the nature of retrospectively developed mapping rules that depend on extrapolation of individual samples to large areas. Mapped values exhibited relatively large changes when mapping rules were modified or when new data were introduced, which is characteristic of models that are over fitted. Recognizing that maps of contaminant distributions are uncertain, EPA evaluated the application of the maps to generate forecasts of correspondence between post-remedial SWAC and RAL.

The effect of uncertainty in the mapping combined with the implicit assumption that each data point represents the average concentration within each Thiessen polygon was evaluated by considering RI and design data from River Mile 10.9. It was found that forecasts based on the Mapping Approach understated the post-remedial SWAC and remedial footprint. To address the possibility that this might have been an anecdotal finding particular to River Mile 10.9, a simulation study was also conducted, which found that the SWAC vs RAL procedure also overstated remedial performance in general, and that the primary factor was falsely classifying areas with contaminated sediments as non-target areas.

As a result of the findings, EPA provides the following points for further discussion:

1) Consider restricting usage of mapped values to estimation of site-wide average concentration, imputation of surface averages as model initial conditions within relatively large areas, and for developing weighted averages within relatively large subareas of the LPRSA that might be used to forecast remedial options.
2) Consider assuring that areas for which averages are to be calculated are spatially contiguous and include multiple replicate samples (i.e. several Thiessen polygons).
3) Consider propagating uncertainty in estimated averages into subsequent calculations that depend on these averages, particularly forecasts of their influence on post-remedial SWAC.
4) More laterally extensive remedial options may be worth evaluating.

# 8 References


AQEA 2013, Lower Passaic River Surface Sediment Concentration Mapping. Memorandum to Robert Law, de maximus, inc. May 17, 2013. From John Connolly, Peter Israelsson, Alyssa Thorvaldsen and George Dang of Anchor QEA, LLC. 236, Broadway, Suite 2. Cambridge, MA.

Burton, A., Altman, D.G., Royston, P, and R.L. Holder. 2006. The design of simulation studies in medical statistics. *Statistics in Medicine.* **25:** 4279-4292.

Connolly J. 2014. *2,3,7,8- TCDD Patterns in Lower Passaic River Sediments.* Available online at Montclair College. Passaic River Symposium VI, Passaic River Institute of Montclair State University, October 9, 2014. http://www.montclair.edu/media/montclairedu/csam/pri/conferencemedia/john-collins-sedminet-2-3-7-8.pdf; Last accessed 4/3/2015.

Cooperating Parties Group 2015. Appendix J Mapping of contaminant concentrations in Lower Passaic River surface sediments. Draft Lower Passaic River Study Area Remedial Investigation Report.

USEPA, 2013. *ProUCL Version 5.0.00 Users Guide. Statistical Software for Environmental Applications for Data Sets with and without Nondetect Observations.* EPA/600/R-07/041 September 2013.

Journel, A.G., and Huijbregts, C.J. 1978. *Mining Geostatistics*. Academic Press. Harcourt Brace and Company, New York.

Kern, J. W., Gbondo-Tugbawa, S.T., Garvey, E.A., Atmadja, J., Prabhu, C. 2009. Mining a rich data set: Evaluation of sediment stability at varying spatial scales on the Lower Passaic River using geostatistical conditional simulation of bathymetric survey data. Proceedings of the 30th Annual Meeting of Society of Environmental Toxicology and Chemistry, New Orleans, LA.

Kern, J.W., Wolfe, J.R. and N. Barabas. 2009b. *Sampling Density for Refinement of Dredge Prism Design in the Lower Fox River, Wsconsin OU2-5*. Proceedings of the 5th International Conference on Remdiation of Contaminated Sediments. New Orleans, LA.

LimnoTech 2013. Observations of Anchor QEA May 17, 2013 Lower Passaic River Surface Sediment Concentration Mapping. Memorandum to Janine McGregor, New Jersey Department of Environmental Protection. September 24, 2013. LimnoTech Inc. 3608 Crest Drive, Manhattan Beach, CA 90266.

Smith, T.M.F. 1983. On the validity of inferences from non-random sample. *Journal of the Royal Statistical Society. Series A* (General), Vol 146, No. 4, pp. 394-403.